THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR BAKHTIAR MOAZZAMI

TITLE OF THESIS A MONTE CARLO STUDY OF THE SMALL SAMPLE PROPERTIES OF SIMULTANEOUS EQUATION ESTIMATORS IN THE PRESENCE OF AUTOCORRELATED DISTURBANCES

DEGREE FOR WHICH THESIS WAS PRESENTED DOCTOR OF PHILOSOPHY

YEAR THIS DEGREE GRANTED SPRING 1984

THE UNIVERSITY OF ALBERTA

A MONTE CARLO STUDY OF THE SMALL SAMPLE PROPERTIES OF SIMULTANEOUS EQUATION ESTIMATORS IN THE PRESENCE OF AUTOCORRELATED DISTURBANCES

by

BAKHTIAR MOAZZAMI

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

ECONOMICS

EDMONTON, ALBERTA

SPRING 1984

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled A MONTE CARLO STUDY OF THE SMALL SAMPLE PROPERTIES OF SIMULTANEOUS EQUATION ESTIMATORS IN THE PRESENCE OF AUTOCORRELATED DISTURBANCES submitted by BAKHTIAR MOAZZAMI in partial fulfilment of the requirements for the degree of DOCTOR OF PHILOSOPHY.

ABSTRACT

Time series data often generate disturbances which are time dependent. The simplest form of this time dependence is first order autocorrelation. Numerous limited information estimators has been proposed for estimation of autocorrelated models. A chronological list of these methods would include Theil(1958), Sargan(1961), Amemiya(1966), Fair(1970,1972), Dhrymes, Berner and Cummins(1974), and Hatanaka(1976). These methods can be distinguished by whether they employ T or T-1 observations, the reduced form they use, the way they estimate the autocorrelation coefficient and whether they allow for the presence of lagged endogenous variables in the equation.

Most of the proposed estimators are asymptotically efficient, but nothing is known about their small sample properties. An econometric practitioner invariably deals with small samples. Therefore, it is of great interest to inquire into the small sample behaviour of these estimators. The purpose of this thesis is to investigate, using Monte Carlo techniques, whether there exist significant differences between the small sample performance of these estimators.

For reasons of cost and relevance, this study has focused on the limited information methods. We have considered almost all the limited information methods proposed in the literature. We have also suggested new methods such as a Generalized Limited Information Maximum

Likelihood Estimator and different ways of interpreting or deriving the existing ones. We have exploited the instrumental variable approach to the Theil estimator to generate modification of Theil's estimator as well as the Brundy and Jorgenson estimators.

This work consists of two major parts. The first part examines the small sample properties of the limited information estimators designed for estimation of autocorrelated models without lagged endogenous variables. The second part studies the small sample properties of dynamic simultaneous autoreggressive model estimators.

We have explored a number of subsidiary issues such as the efficiency gain of employing the first observation, utilization of different reduced forms, alternative specification of the exogenous variables and different ways of estimating the autocorrelation coefficient. The relative importance of the first observation in the single equation context was investigated recently by Maeshiro(1976,1979), Park and Mitchell(1980), and Beach and McKinnon(1980). They found instances in which the omission of the first observation caused by the autoregressive transformation resulted in substantial loss of efficiency. We have examined whether their results hold in the simultaneous equation context.

The major finding of this study is that the Theil Generalized Two Stage Least Squared estimator, which has been completely ignored in the applied literature,

unambiguously dominates all other estimators in the static model. In the case of the dynamic model Theil's estimator, Dhrymes' Convergent Two Stage Least Square estimator and Hatanaka's residual adjusted estimators dominated all other estimators. Fair's estimator, which is the most commonly used method, proved to be extremely inefficient.

ACKNOWLEDGEMENT

My sincere thanks are due to Dr. A. Buse. I have been fortunate to have him as my supervisor. He provided generous encouragement, sympathetic advice and acute criticism throughout the various stages of this study, without which the thesis could not have been completed. I also wish to acknowledge the invaluable guidance and assistance provided by Drs. B. Von Hohenbalken and M. Hellsten, the other members of my supervisory committee. In addition, I am indebted to Drs. G. Fisher and D. Jobson, my external examiners, for their valuable suggestions and comments. My appreciation also goes to Dr. S. Lewis and A. Sharpe for their help in computer programming throughout this research.

Many others have assisted me in the process of writing this thesis. My special thanks go to my friend and colleague Dr. S. Manouchehri with whom I discussed parts of this thesis. I also want to express my gratitude to my friends C. Nunn, G. Grim, A. Sepehri and M. Nasseri for the intelectual and moral support they provided me during the past few years. I would like to thank Maimanat and Yadollah for their friendship and encouragement. Finally, I owe an everlasting debt to my parents and to Ardeshir and Dara, my brothers.

Table of Contents

List of Tables

I. INTRODUCTION

This chapter reviews some of the issues and problems involved in estimation of simultaneous equation models characterized by first order autocorrelation. To provide a suitable framework for the subsequent discussion single equation models are first reviewed. Then the case of simultaneous equation systems is introduced. Finally the issues and problems of this thesis will be presented and the plan of the work will be outlined.

A. The Single Equation Model

Problems associated with the presence of autocorrelation in the context of the single equation model have been extensively studied.[1] Many of the methods appropriate for estimation of models with autocorrelated errors are two step procedures, i.e., transformation of the autocorrelated model to one free of autocorrelation and estimation of the resultant equation using Ordinary Least Square (OLS). However, these methods differ in their choice of transformation matrix and also in the way they estimate the autocorrelation coefficient.

Generally, they can be regarded as special cases of the Maximum Likelihood (ML) procedure developed by Beach and Mackinnon(1978). Consider the standard linear regression model with a first order autoregressive error structure defined by

----

[1]For a recent summary of much of this work see Judge, Griffiths, Hill and Lee(1980).

$$Y = X\beta + U$$

$$U_t = r\,U_{t-1} + E_t \qquad t = 1,\ldots,T \tag{1}$$

$$E_t \sim IN(0,\sigma^2)$$

where Y is a T.1 vector of observations on the dependent variable, X is T.K matrix of observations on the exogenous variable, U and E are a T.1 vectors of the error terms, β is a K.1 vector of coefficients, and r is a coefficient of autocorrelation.

The concentrated log-likelihood function corresponding to the model (1) is

$$L = \text{const} + 1/2\log(1-r^2) - T/2\log[(1-r^2)(Y_1-X_1'\beta)^2 + \sum_{2}^{T}(Y_t-X_t'\beta-rY_{t-1}+rX_{t-1}'\beta)^2] \tag{2}$$

where X_i' is the ith row of the X matrix.

The first order conditions with respect to β and r yield the maximum likelihood estimators of those coefficients as

$$\hat{\beta} = (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y}$$

where

$$\dot{X} = PX \;;\; \dot{Y} = PY$$

and P is the T.T Prais-Winsten(1954) transformation matrix

$$P = \begin{pmatrix} \sqrt{(1-r^2)} & 0 & \ldots & & \\ -r & 1 & 0 & \ldots & \\ 0 & -r & 1 & 0 & \ldots \\ & & & \ldots & \\ & & \ldots & -r & 1 \end{pmatrix}$$

and the solution for the autocorrelation coefficient, r, is cubic with one real root in the interval (-1,1).

Generalized Least Squares(GLS) estimation of the model (1) is equivalent to the maximization of the log-likelihood (2) while omitting the first term (the Jacobian of the transformation), i.e., the term which ensures that the stationarity condition holds. The GLS estimators of β and r are then given by

$$\hat{\beta} = (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y}$$

$$\hat{r} = \sum_{2}^{T}\hat{U}_t\hat{U}_{t-1} / \sum_{3}^{T} \hat{U}^2_{t-1}$$

where $\hat{U}_t$ is the estimated vector of the error terms, U_t.

It can be seen that GLS procedure yields analytically the same formula for estimation of the coefficient β as the ML method. The only difference is that the GLS procedure estimates the coefficient of autocorrelation from the Prais-Winsten formula whereas in the ML method it is being estimated via the solution to a cubic equation.

The third alternative, which has until recently been the most commonly used method is the Cochrane-Orcutt(1949) procedure. This method is equivalent to maximizing the log-likelihood function (2) with respect to β and r while omitting the terms $\log(1-r^2)$ and $(1-r^2)(Y_1-X_1'\beta)^2$. Estimators of β and r obtained from the Cochrane-Orcutt procedure are given by

$$\hat{\beta} = (\overline{X}'\overline{X})^{-1}\overline{X}'\overline{Y}$$

$$\hat{r} = \sum_{2}^{T}\hat{U}_t \hat{U}_{t-1} / \sum_{2}^{T} \hat{U}^2_{t-1}$$

where $\overline{X} = QX$, $\overline{Y} = QY$,

and Q is the Cochrane-Orcutt(CORC) transformation matrix which is the same as the matrix P without its first row. In

this procedure the coefficient of autocorrelation is estimated by regressing the current residual on its one period lag. The reason for the popularity of this method is its simplicity since it is nothing more than applying OLS to the transformed equation

$$Y_t - rY_{t-1} = (X_t - rX_{t-1})'\beta + E_t \qquad t = 2,\ldots,T$$

Note, however, that unlike the ML and GLS estimators, this method uses only T-1 of the T observations. It should also be noted that it is customary to iterate the GLS and CORC procedures up to a convergence criteria.

All three procedures are asymptotically equivalent and hence share the ML properties of consistency, efficiency and normality. That is to say, the effect of the first observation diminishes as the sample size increases. But the omission of the first observation may have serious effects on the small sample performance of the estimators. We now briefly review the literature pertaining to this issue.

The first major work dealing with the effects of autocorrelation on the small sample properties of the estimators in the single equation context was undertaken by Rao and Griliches(1969). They employed a standard linear regression model as defined by (1). They conducted their experiment with a sample of 20 observations using an explanatory variable generated by a first order Markov process, i.e., $X_t = \lambda X_{t-1} + V_t$, with λ ranging from -0.8 to 0.99. They considered several two-step regression methods designed for estimation of model (1) and compared them with

OLS. Their experiment suggested that[2]

> there is a significant gain in efficiency to be had from using two stage estimation procedures for moderate and high levels of serial correlation in the residuals (r > 0.3) and very little loss from using such methods even when the true r is small.

They also reported some gain in efficiency for those estimators that retain the first observation over the ones which ignore it.

The results of the Rao and Griliches study were criticised in a series of articles by Maeshiro(1976,1979) who showed that(1976,p.500)

> ...contrary to its intended purpose, the Cochrane-Orcutt transformation reduces rather than increases the efficiency of estimators in many cases...thus the various estimation methods proposed by Cochrane and Orcutt(1949) and the method proposed by Durbin(1960,pp.150,153) are dubious. In fact, contrary to the expectations of Cochrane and Orcutt and Durbin, ordinary least squares estimators may work better for many cases of trended independent variables. This conclusion also leads us to cast serious doubt about the validity of the findings of various Monte Carlo studies that do not use trended independent variables as a guide in the choice of estimators when an independent variable contains a trend.

Maeshiro(1979) further showed that the gain in efficiency of those estimators that utilize the first observation is much larger than what was suggested by Rao and Griliches, if the independent variable contains a moderate trend.

Maeshiro's results were analytically confirmed by Chipman(1979). He provided a formula for the minimum lower bound of efficiency of CORC and OLS estimates when the explanatory variable is trended. He showed that in the case

----

[2] Rao and Griliches, (1969), pp.267-68.

where the explanatory variable is a simple linear trend, "the Cochrane-Orcutt procedure is only 71 percent as efficient as ordinary least squares".[3]

A heuristic explanation of the observed pattern of the poor performance of CORC estimator vis-a-vis OLS and GLS is given by Maeshiro(1979,pp.259-60) as follows

> The reason why a substantial gain in efficiency can be expected by retaining the first observation in the case of trended independent variables but not necessarily in other cases is twofold. First, the autoregressive transformation (with positive r) of a trended variable tends to result in a new independent variable that takes more similar values and hence possesses less variability than the original independent variable. Second, the weighted first observation $(\sqrt{1-r^2})X_1$, on the other hand, tends to take a value that is quite different from the values of the new independent variable created by the autoregressive transformation. As a result, the marginal contribution of the weighted observation to the increase in the variability of the new independent variable and, therefore, to the increase in the efficiency of the estimator, can be substantial....For nontrended independent variables, however, the autoregressive transformation of a variable does not necessarily reduce (in fact, can increase) variability and the weighted first observation is not necessarily very different from the values of the autoregressive transformed variable.

Park and Mitchell(1980) also confirmed the findings of Maeshiro. They also examined the performance of different estimators in hypothesis testing. They found that(1980,p.185), [4]

> none of the feasible estimators performs well in hypothesis testing; all seriously understimate standard errors, making estimated coefficients appear to be much more significant than they actually are.

----

[3]Chipman,(1979), p.117.
[4] Park and Mitchell, (1980), p.185.

Beach and Mackinnon(1978) also compared the small sample performance of the Cochrane-Orcutt and full information maximum likelihood estimators (FIML). They found that FIML which incorporates all the observations always performed better than the CORC procedure and this difference was specially noticeable when the exogenous variable was trended.

Taylor(1981), using analytical approximations, showed that performance of the single equation estimators is sensitive to the way the explanatory variable is specified. He showed that the value of the first observation becomes extremely important if the explanatory variable is non-stochastic and trended and its trend coefficient happens to be close to the autocorrelation coefficient of the error term.[5]

Finally, Doran and Grifitts(1982) conducted experiments using the seemingly unrelated regression model with first order autoregressive disturbances. They were particularly interested in the relative efficiency of using the first observation. They found that there does not exist a significant difference between the estimators that utilize T and the ones that employ T-1 observations.

---

[5]We shall review his findings in more detail later.

B. Simultaneous Equation Models

Economic models usually involve a set of interrelated relationships aimed at explanation of certain economic variables. In general, they can be compactly written as

$$Y\,B + X\,\Gamma = U \tag{3}$$

where Y is the T.G matrix of observations on the endogenous variables, X is the T.K matrix of observations on the predetermined variables, B is the G.G matrix of structural coefficients associated with the endogenous variables, Γ is the K.G matrix of coefficients of the predetermined variables, and U is the T.G matrix of disturbances.

For estimation of the system (3), the following assumptions about the stochastic disturbances are usually made[6]

$$U_{ti} \sim N(0,\sigma^2), \qquad i = 1,\ldots,G$$

$$E(\,U_{ti}\;U_{sj}\,) = 0 \qquad t,s = 1,\ldots,T$$

$$t \neq s$$

$$E(\,U_{ti}\;U_{tj}\,) = \sigma_{ij} \qquad i,j = 1,\ldots,G$$

In matrix notation these assumptions become

$$U_t \sim N(0,\Omega) \tag{4}$$

$$E(\,U_t'\;U_s\,) = 0 \qquad t \neq s \tag{5}$$

where U_i is the ith row of the U matrix, and

$$\Omega = \begin{pmatrix} \sigma_{11},\ldots, & \sigma_{1G} \\ & \\ \sigma_{G1},\ldots, & \sigma_{GG} \end{pmatrix} \tag{6}$$

----

[6] We note here that in the subsequent development E has been used to signify an error term. However, in some instances we have used E to denote expectation. The distinction between these usages is, however, clear from the context.

Given equations (4), (5), and (6) and the assumption of identifiability, system (3) can be consistently estimated using full or limited information methods. Methods appropriate for the estimation of this type of model are the ones which take into account the simultaneity created by the presence of the current endogenous variables among the explanatory variables of each structural equation.

However, one of the crucial assumptions made about the properties of the stochastic disturbances is not usually satisfied. As in the single equation case, time series data invariably generate disturbances which are time dependent. The simplest form of this time dependence is a first order vector autoregressive process which replaces assumption (5) by

$$U_t = U_{t-1} R + E_t \tag{7}$$

where E_t is a random vector of G components, i.e., $E_t' = (\epsilon_{1t}, \ldots, \epsilon_{Gt})$, $t=1,\ldots,T$; the E_t for any two different time periods being independently normally distributed with mean zero and a constant variance-covariance Θ, and R is a G.G matrix of autocorrelation coefficients.

Following the practice in Amemiya (1961), Fair(1972), and Dhrymes(1974), we assume that R is a diagonal matrix, i.e.

$$R=\mathrm{diag}(r_1, r_2, \ldots, r_G)$$

where

$$-1 < r_i < +1 \qquad i = 1,\ldots,G$$

It is important to note that the assumption of diagonality of R does not imply independence of the error terms across equations. One can easily show that

$$E(U_t'U_t) = \Omega = R'\Omega R + \Theta$$

The presence of autocorrelation poses additional problems for the estimation of the simultaneous equation system formulated in (3) and (7). In general, the application of the classical linear simultaneous equation techniques which ignore the autocorrelation results in inefficient, and in some cases(i.e., lagged endogenous variables) inconsistent estimates of the structural coefficients.

Techniques appropriate for estimation of systems characterized by autocorrelation are not adequately addressed in the textbooks. Only a limited number of textbooks have tackled this problem and even some of those have treated it inadequately.[7] However, there has been a series of articles in the journal literature proposing different methods for estimation of medels with autocorrelation. A systematic review of these methods will be presented in later chapters. However, some of the general problems and issues involved in estimation of the simultaneous equation models with autocorrelation is discussed in the next section. Then, the structure of this thesis and the issues that will be studied are presented.

----

[7]See Kmenta(1971), Pindyck and Rubinfeld(1981), Klein(1974), Kelejian and Oates(1981) and Stewart and Wallis(1981).

C. Estimation of Simultaneous Equation System in the Presence of First Order Autocorrelation

The model described by (3) and (7), namely

$$Y B + X \Gamma = U \tag{8}$$

$$U_t = U_{t-1} R + E_t \tag{9}$$

can be written as a restricted transformed structure by eliminating (9) to obtain

$$Y B + X \Gamma - Y_{-1} B R - X_{-1} \Gamma R = E \tag{10}$$

The system represented in (8) and (9) can be estimated using system or limited information techniques. Estimating via limited information methods ignores three sources of information about all equations except the one which is being estimated, namely,

1- Any across equation over-identifying restrictions.

2- Across equation covariance structures.

3- The restrictions caused by the autoregressive transformation in (10).

However, on the credit side, the limited information methods are simpler to implement and, moreover, they prevent any specification error ocurring in one equation from affecting the estimation of the other equations.

In practice limited information methods are most often used and these are the set of estimators examined in this thesis. Unfortunately, there is no unified systematic treatment of the methods for estimation of system (8) and (9) in the literature. A large number of alternative estimators has been proposed. These estimators, often

confusing, can be distinguished by whether they use T or T-1 observations, by the structure of the reduced form they employ and by the way they estimate the autocorrelation coefficient. We now consider these issues.

In practice the coefficient of autocorrelation is never known. Usually, a consistent estimate of it based on the estimated residuals from a first-stage consistent estimator is employed. In the single equation context, there are a number of alternative methods proposed for estimation of the autocorrelation coefficient. We shall review them here and expect that their properties carry over to the simultaneous system as well. Basically, these alternative estimators are obtained by maximizing the log likelihood (2) with respect to r conditional on a given value of the vector β. If we ignore the Jacobian of the transformation in (2), the above procedure amounts to minimizing the sum of squared residuals given in the last term of (2) with respect to r for a given value of β. This means that for those estimators which use T-1 observations, r will be estimated as the value which minimizes the sum of squared T-1 residuals as

$$SR_1 = \sum_{2}^{T} \hat{E}_t^2 = \sum_{2}^{T} (\hat{U}_t - r\hat{U}_{t-1})^2$$

since $U_t = r\, U_{t-1} + E_t$

To minimize SR_1 with respect to r, we have

$$\partial SR_1 / \partial r = 0$$

which yields

$$\hat{r}_1 = \sum_{2}^{T} \hat{U}_t \hat{U}_{t-1} \Big/ \sum_{2}^{T} \hat{U}^2_{t-1}$$

which is the familiar Cochrane-Orcutt formula.

The same process applies to the estimators which utilize all T observations. The r which minimizes the sum of squared T residuals can be estimated by partially differentiating SR_2 (defined below) with respect to r and setting the first order condition equal to zero

$$SR_2 = \sum_{1}^{T} \hat{E}_t^2 = (1-r^2)\hat{U}_1^2 + \sum_{2}^{T}(\hat{U}_t - r\hat{U}_{t-1})^2$$

$$\partial SR_2 / \partial r = 0$$

yields

$$\hat{r}_2 = \sum_{2}^{T} \hat{U}_t \hat{U}_{t-1} / \sum_{3}^{T} \hat{U}^2_{t-1}$$

which is the Prais-Winsten formula for estimating the autocorrelation coefficient.

Obviously $\hat{r}_2$ tends to be greater than $\hat{r}_1$. Theil(1971) proposed an alternative estimator that incorporates a degrees of freedom correction which would result in a smaller estimate of r than $\hat{r}_2$ and $\hat{r}_1$. However, empirical research (see for instance the essay by Hildreth and Dent in Sellekaerts(1974)) shows that $\hat{r}_2$ and $\hat{r}_1$ tend to have a downward bias as they stand. In the light of this fact, use of the Theil estimator does not seem appropriate. In the present work, we will follow Park and Mitchell(1980) and use the Prais-Winsten formula for estimators that use all T observations and the CORC method for those which utilize T-1 observations. We emphasize again that these estimators are derived from the single equation case and it is a matter for empirical determination whether these differences are important in the simultaneous equation case.

With regard to the employment of the reduced form structure, there are two possibilities. The reduced form system can be written as

$$Y = -X \Gamma B^{-1} + U B^{-1} \tag{11}$$

or in the form called "the augmented reduced form", that is

$$Y = -X \Gamma B^{-1} + Y_{-1} B R B^{-1} + X_{-1} \Gamma R B^{-1} + E B^{-1} \tag{12}$$

which is obtained by substituting for U in (3).

In the absence of lagged endogenous variables, the reduced form (11) can be utilized to provide consistent estimates of the endogenous variables appearing on the right hand side of any structural equation. But if the system contains lagged endogenous variables, the reduced form (12) should be used. However, the distinction has not always been clearly made in the literature.

D. Monte Carlo Evidence

Most of the estimators that have been proposed have yielded up large sample properties, but the small sample properties of these estimators has not been considered. This question can be answered by theoretical analysis or by Monte Carlo experiments. Even though there has been some attempt to arrive at the theoretical derivations of the small sample distributions of different estimators[8] , their derivation is

----

[8]Nagar(1959), Basmann(1961), Richardson(1968), Sawa(1969), Sargan and Mikhail(1971), Kadane(1971), Phillips(1980), and Marino(1982) among others have derived the finite sample distributions and the first few moments (where they exist) for K-class estimators under the assumptions of non-stochastic exogenous variables in the absence of autocorrelation.

often time consuming, difficult and sometimes impossible. The difficulties involved in the theoretical derivation are the main reason for the attractiveness of the Monte Carlo methods.[9] Monte Carlo experiments tend to have lower labor cost and their comparative advantage increases with the complexity of the problem.[10]

Results of works by Phillips(1977) and Maasoumi(1980) among others suggest that the discrepancies between asymptotic and finite sample behavior of the estimators are parameter dependent and we cannot, without qualification, postulate the asymptotic behavior of the estimators from the small sample results. However, as Hendry(1973, 1974) has suggested we can use the asymptotic properties as a guide to the small sample performance of different estimators.

After an extensive review of the literature, it was found that there exist only a few studies which have tried to investigate the small sample properties of the simultaneous equation estimators in the presence of autocorrelation. Schink and Chiu(1966) examined the effects of autocorrelation on the performance of the ordinary least squares, limited information single equation (LISE) and two stage least squares estimators. They concluded that autocorrelation had no significant effect on the performance of LISE and 2SLS, while it affects negatively the OLS estimator. Hurd(1972) examined the performance of OLS, CORC,

----

[9]See Summers,1965.
[10]For a survey of these studies see Johnston(1972), Sowey (1973).

2SLS and a modified version of 2SLS which in addition to simultaneity also corrected for the autocorrelation.[11] He found that techniques which correct only the autocorrelation perform much better than those which correct for the simultaneity. Goldfeld and Quandt(1972) examined the small sample performance of OLS, 2SLS and various versions of FIML. They found that FIML when it takes into account both autocorrelation and simultaneity performs better than OLS, 2SLS and different versions of FIML that ignore either of the simultaneity or autocorrelation problems.

Hendry and Harrison(1974) investigated the main sources of small sample inconsistencies of OLS and 2SLS when they are applied to an structural equation with autocorrelation. Hendry and Srba(1977) investigated the performance of the OLS, instrumental variable estimator(IV) and their generalization which allowed for the presence of autocorrelation (denoted as ALS and AIV), and ordinary 2SLS estimators using a dynamic autoregressive model. They concluded that asymptotic distribution theory can be useful in assessing the small sample behavior of different estimators. Finally, Wang and Fuller(1982), using a dynamic autoregressive model, compared the performance of a autoregressive two-stage least squares estimator with two full information estimators suggested by Wang and Fuller(1975) and independently derived by Hatanaka(1976).

----

[11]It is worth noting that the modified 2SLS estimator considered by Hurd is equivalent to the modified Theil estimator analysed in chapters II and III of this thesis.

All three estimators shared some features of the two-step Gauss-Newton procedure and of Aitken generalized least squares. They found that(1982, p.140) "Monte Carlo results for the autoregressive estimators are generally consistent with the large sample properties of those estimators."

However, the scope of these works on the small sample properties of the simultaneous equation estimators in the presence of autocorrelation is very limited. Nothing is known about the small sample properties of most of the estimators such as Theil's Generalized Two Stage Least Squares(G2SLS), Brundy and Jorgenson's instrumental variable estimator, Amemiya's autoregressive Limited Information Maximum Likelihood (ALIML) estimator, Fair's estimator, Hatanaka's two step methods, and Dhrymes' two step procedures. The relative lack of knowledge in this important area of econometrics is the primary reason for undertaking this thesis. Our objective is to investigate the small sample performance of all major simultaneous equation estimators designed for estimation of the models characterized by autocorrelation.

To pursue this task, we shall concentrate on the limited information methods. This choice is primarily due to cost considerations but the fact that these methods are of more practical importance than the full information ones is also a major justification of our approach. We have divided our enquiry into two major parts. The first part investigates the small sample properties of the limited

information methods of estimation in the absence of lagged endogenous variables. The primary focus of this part is the investigation of the effects of utilizing T or T-1 observations, employing different reduced forms, and using different methods for the estimation of the autocorrelation coefficient. This part also investigates the effects of different specifications of the exogenous variables on the small sample properties of the estimators. The second part of this work studies the small sample properties of the limited information estimators in the presence of lagged endogenous variables.

II. LIMITED INFORMATION METHODS OF ESTIMATION IN THE ABSENCE OF LAGGED ENDOGENOUS VARIABLES

This chapter presents the limited information methods of estimation of autocorrelated models without lagged endogenous variables. These estimators can be categorized in a variety of ways. They can be classified according to whether they employ T or T-1 observations. The number of observations utilized has a direct bearing on the type of reduced form that they can employ and the way they estimate the autocorrelation coefficient. They can be distiguished on the basis of the instruments or the reduced form that they employ. For the purpose of exposition we classify them into the following two categories: (i) Those that are generated by the Limited Information Maximum Likelihood (LIML) approach and (ii) those that follow the Theil Generalized Two Stage Least Square (G2SLS) approach. The emphasis in this chapter will be on the computational aspects of the estimators. In the next chapter the asymptotic properties of the estimators will be considered paying particular attention to the question of asymptotic efficiency.

A. The Variants of the LIML Approach

Autoregressive Limited Information Maximum Likelihood (ALIML)

Without loss of generality, we focus on the estimation of the first equation of the system considered in Chapter I. We can write this equation (after the normalization $\beta_{11}=1$)

as

$$y_1 = Y_1 \beta_1 + X_1 \gamma_1 + u_1 \tag{1}$$

where

$$u_1 = r u_{1,-1} + \epsilon_1 , \tag{2}$$

y_1,Y_1,X_1 and u_1 are sub-matrices of Y, X, and U, and where y_1 is a T.1 vector of values of the dependent variable, Y_1 is an T.G_1 matrix of endogenous variables other than the first one included in the first equation, X_1 is a T.K_1 matrix of exogenous variables included in the first equation, u_1 and ϵ_1 are T.1 vectors of disturbance terms, r is the (1,1) element in R, and β_1 and γ_1 are G_1.1 and K_1.1 vectors of coefficients corresponding to the relevant elements of B and Γ, respectively.

Substituting (2) in (1) yields

$$y_1 - r y_{1,-1} = (Y_1 - r Y_{1,-1})\beta_1 + (X_1 - r X_{1,-1})\gamma_1 + \epsilon_1 \tag{3}$$

Using the notation $\overline{Y}_i = Y_i - r Y_{i,-1}$, etc, (3) becomes

$$\overline{y}_1 = \overline{Y}_1 \beta_1 + \overline{X}_1 \gamma_1 + \epsilon_1 \tag{4}$$

or

$$\overline{y}_1 = \overline{Z}_1 \delta_1 + \epsilon_1 \tag{5}$$

The problem of estimating (5) was first addressed by Sargan(1961). He proposed a Limited Information Maximum likelihood method for the estimation of this equation. His estimator was later modified by Amemiya(1966) to arrive at some two step alternatives. The concentrated likelihood function corresponding to the system (1) is given by

$$L_1 = K_1 + 1/2 \log(B_1' W B_1) - 1/2 \log(\epsilon_1' \epsilon) \qquad (6)$$

where

$$W = Y_1^{*\prime}(I - \Phi(\Phi'\Phi)^{-1}\Phi')Y_1^*$$

$$\Phi = (X \;, Y_{-1}, \; X_{-1})$$

$$Y_1^* = (y_1 \;, \; Y_1)$$

where B_1 is the first row of the B matrix.

Using Sargan's estimator(ALIML), the estimates of the coefficients of the system (1) can be obtained by partially differentiating (6) with respect to β_1, γ_1, and r and solving the resultant first order conditions. The solution for β_1 and γ_1 is

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{pmatrix} = \begin{pmatrix} \bar{Y}_1'\bar{Y}_1 - \lambda \bar{Y}_1'(I - \Phi(\Phi'\Phi)^{-1}\Phi')\bar{Y}_1 & \bar{Y}_1'\bar{X}_1 \\ \bar{X}_1'\bar{Y}_1 & \bar{X}_1'\bar{X}_1 \end{pmatrix}^{-1}$$

$$\begin{pmatrix} \bar{Y}_1 - \lambda \bar{Y}_1(I - \Phi'(\Phi'\Phi)^{-1}\Phi') \\ \bar{X}_1' \end{pmatrix} \bar{y}_1 \qquad (7)$$

where λ is the smallest root of the determinantal equation

$$|W_1 - \lambda W| = 0$$

and

$$W_1 = \bar{Y}_1^{*\prime}(I - \bar{X}_1(\bar{X}_1'\bar{X}_1)^{-1}\bar{X}_1')\bar{Y}_1^*$$

The autocorrelation coefficient is estimated as

$$\hat{r} = (\hat{u}_t'\hat{u}_{t-1}) / \hat{u}_{t-1}^2 \qquad (8)$$

where

$$\hat{u} = y_1 - \hat{y}_1$$

A modified version of the above estimator called Sargan 2SLS (S2SLS) was proposed by Amemiya(1966). S2SLS is obtained by setting λ=1 which gives

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{pmatrix} = \begin{pmatrix} \bar{Y}_1'\Phi(\Phi'\Phi)^{-1}\Phi'\bar{Y}_1 & \bar{Y}_1'\bar{X}_1 \\ \bar{X}_1'\bar{Y}_1 & \bar{X}_1'\bar{X}_1 \end{pmatrix}^{-1} \begin{pmatrix} \bar{Y}_1'\Phi(\Phi'\Phi)^{-1}\Phi' \\ \bar{X}_1' \end{pmatrix} \bar{y}_1$$

The S2SLS estimator of r is identical to that of ALIML which is given in (8). S2SLS is related to the ALIML estimator in the same way that 2SLS is related to LIML. The S2SLS estimator can be regarded as an instrumental variable estimator which uses all the X's, and one period lags of the X's and the Y's as instruments for the right hand side endogenous variables.[12] These instruments are all the variables entering the augmented reduced form for Y_1(Chapter I, equation (12)).

The main drawback of Sargan's method is the large number of instruments it uses. Even in a moderate size model, there is a possibility that the number of instruments is greater than the number of observations. Moreover, the presence of all the X's and their lagged values in the list of instruments can cause severe multicollinearity that in turn will result in a poor estimates.[13] In the absence of lagged endogenous variables a consistent estimate of Y_1 can be obtained from the ordinary reduced form (Chapter I, equation (11)). This can clearly reduce the number of instruments in the initial stage and thus may decrease the

---

[12]See Sargan(1959, 1961).

[13]Poor in the sense that the fitted values of Y, even though consistent, will have relatively large variances. Whether this affects the small sample properties of the estimator or not is a question that will be empirically investigated later.

small sample bias of the estimator.[14] However, its effect on the small sample performance of the estimators must be investigated.

To tackle the potential degrees of freedom problem, especially for large models, Fair suggested a number of alternative estimators which use fewer instruments than S2SLS. In the following we shall discuss the motivation and the steps required in estimation of his most clearly enunciated procedure which is also incorporated in the TSP package.[15] The asymptotic properties of this estimator will be discussed in the next chapter.

Fair Estimator

Nagar(1959) showed that the small sample bias of the K-class estimators is positively related to the number of predetermined (instrumental) variables in excess of the number of coefficients to be estimated. Using this theorem Fair(1970) showed that the S2SLS which uses instruments in excess of the ones which appear in the reduced form of the estimated equation increases the small sample bias, to the order T^{-1}, of that estimator.[16]

---

[14] See Nagar(1959). We shall discuss this point shortly.
[15] Note that the Fair estimator in TSP uses the Cubic formula developed by Beach and Mackinnon to calculate the autocorrelation coefficient. However, as we shall demonstrate later, this does not have a significant effect on the estimated coefficients.
[16] Fair showed that even when the augmented reduced form is being used, the number of instruments employed in S2SLS can be drastically reduced. He (1970, p.511) demostrated that the explanatory variables that are excluded (X^x) from the structural equation under consideration, appear in the augmented reduced form as ($X_i{}^x - r_i X_{i,-1}{}^x$). This means that

He proposed an alternative estimator[17] which is asymptotically less efficient than S2SLS but uses fewer instruments and thus may have smaller small sample bias.

The structure of Fair's estimator is identical to S2SLS except that it uses a subset of the instruments used by S2SLS, i.e., $X_m=(X_1,X_{1,-1},y_{1,-1},Y_{1,-1})$ where X_1, y_1, and Y_1 are defined above.

Given the autocorrelation coefficient r, we can write the equation under consideration as

$$y_1-r\ y_{1,-1} = (Y_1-r\ Y_{1,-1})\beta_1 + (X_1-r\ X_{1,-1})\gamma_1 + \epsilon_1$$

or

$$\overline{y}_1 = \overline{Y}_1\beta_1 + \overline{X}_1\gamma_1 + \epsilon_1 \qquad (10)$$

The set of instruments used by S2SLS was $X_s= (X, X_{-1}, Y_{-1})$ which are the variables appearing in the augmented reduced form. Let

----

[16](cont'd)X_i^x and its one period lag do not need to appear as two independent instruments; rather their transformed form appears as one instrument in the set of explanatory variables describing the endogenous variables appearing on the right hand side of the equation under consideration. Using the transformed variables can greatly reduce the number of instruments used for the first stage estimation in S2SLS without any loss of efficiency (Fair,1970, pp.510, 511). However, this method which uses the transformed excluded exogenous variables as instruments requires the knowledge of all the structural autocorrelation coefficients. This in turn requires the estimation of a complete system prior to the estimation of the structural equation under consideration which would not be practical especially in large models. Therefore, in this study we have only considered Fair's most widely known method (outlined in this section) which is also used in TSP package with a slight modification.

[17] Note that Fair's estimator was designed in the first instance to deal with the lagged endogenous case. We are including it here because Fair's procedure is widely used for static autoregressive model estimation.

$$X_m = X_s(S_1, S_2, S_3) = (X_1, X_{1,-1}, y_{1,-1}\ Y_{1,-1})$$

where S_1, S_2, and S_3 are appropriate selection matrices which generate the minimum set of instruments needed for the Fair estimator.

Fair replaces $\overline{Y}_1 = Y_1 - r\ Y_{1,-1}$ in (10) by $\hat{Y}_1 - r\ Y_{1,-1}$ where $\hat{Y}_1$ is obtained from a regression of Y_1 on X_m, i.e.,

$$\hat{Y}_1 = X_m(X_m'X_m)^{-1}X_m'Y_1$$

so that

$$\hat{Y}_1 - rY_{1,-1} = X_m(X_m'X_m)^{-1}X_m'Y_1 - rY_{1,-1} \qquad (11)$$

The right hand side of (11) can be written, using the notation $Y_{1,-1} = X_m\ S_{3,-1}$, as

$$\hat{Y}_1 - rY_{1,-1} = X_m(X_m'X_m)^{-1}X_m'Y_1 - rX_m(X_m'X_m)^{-1}X_m'X_mS_{3,-1}$$
$$= X_m(X_m'X_m)^{-1}X_m'(Y_1 - rY_{1,-1})$$

therefore

$$\hat{Y}_1 - rY_{1,-1} = X_m(X_m'X_m)^{-1}X_m'\overline{Y}_1 = \hat{\overline{Y}}_{1m} \qquad (12)$$

Using (12) in place of $\overline{Y}_1$ in (10) and applying OLS gives

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{pmatrix} = \begin{pmatrix} \hat{\overline{Y}}_{1m}'\hat{\overline{Y}}_{1m} & \hat{\overline{Y}}_{1m}'\overline{X}_1 \\ \overline{X}_1'\hat{\overline{Y}}_{1m} & \overline{X}_1'\overline{X}_1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \hat{\overline{Y}}_{1m}'\overline{y}_1 \\ \overline{X}_1'\overline{y}_1 \end{pmatrix} \qquad (13)$$

which can be further simplified as

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{pmatrix} = \begin{pmatrix} \overline{Y}_1'X_m(X_m'X_m)^{-1}X_m'\overline{Y}_1 & \overline{Y}_1'\overline{X}_1 \\ \overline{X}_1'\overline{Y}_1 & \overline{X}_1'\overline{X}_1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \overline{Y}_1'X_m(X_m'X_m)^{-1}X_m'\overline{y}_1 \\ \overline{X}_1'\overline{y}_1 \end{pmatrix} \qquad (14)$$

Which is exactly the same as S2SLS with X_s replaced by X_m. Equation (14) shows that the Fair and S2SLS are identical in structure. This is of course what one would expect since

S2SLS regresses Y_1 on X_s whereas Fair regresses Y_1 on X_m, i.e., Fair chooses a sub-set of instruments from X_s and because it chooses a sub-set of instruments it will also be inefficient relative to S2SLS.

Fair Brundy and Jorgenson Estimator

The primary purpose of the Fair estimator is to overcome the potential degrees of freedom problem in the estimation of the reduced form by using a sub-set of the variables appearing in the augmented reduced form as instruments. An alternative approach to solve this problem in the estimation of simultaneous equation systems was proposed by Brundy and Jorgenson(1971) and was modified by Fair(1972), to take into account the autoregressive properties of the error terms. This estimator does not require reduced form estimation at the initial stage. At the first stage it takes into account the restrictions imposed on the structural equation of the system other than the one under consideration. Fair's version of Brundy-Jorgenson estimator is constructed as follows:

Write the augmented reduced form

$$Y = -X \Gamma B^{-1} + Y_{-1} B R B^{-1} + X_{-1} \Gamma R B^{-1} + E B^{-1} \tag{15}$$

as

$$Y = \Phi \Pi^* + V \tag{16}$$

where $\Phi = (X \;, Y_{-1}, X_{-1})$, $V = E B^{-1}$ and Π^* is partitioned according to Φ. Brundy and Jorgenson showed that[18]

---

[18]Actually the Brundy and Jorgenson proof was for the case where there was no autocorrelation. By analogy, Fair has

any set of instrumental variables based on a consistent estimate of $\Pi*$ will result in asymptotically efficient estimates of the structural parameter, β_1, and γ_1. $\Pi*$ is a function of the B, Γ, and R matrices. Therefore, any consistent estimate of the latter can be used to form $\Pi*$. In the absence of lagged endogenous variables, application of 2SLS to each structural equation provides consistent estimates of the structural as well as the autocorrelation coefficients. These estimates can be used to form consistent estimates of the B, Γ, and R matrices to give $\hat{\Pi}*$ which in turn is used to generate $\hat{Y} = \Phi\ \hat{\Pi}*$. In the model

$$y_1 - r\ y_{1,-1} = (Y_1 - r\ Y_{1,-1})\beta_1 + (X_1 - r\ X_{1,-1})\gamma_1 + \epsilon_1 \quad (17)$$

define matrices $\bar{y}_1$ and $\bar{Z}_1$ and δ as[19]

$$\bar{Z}_1 = [(Y_1 - \hat{r}\ Y_{1,-1})\ ,\ (X_1 - \hat{r}\ X_{1,-1})]$$

$$\bar{y}_1 = y_1 - \hat{r}\ y_{1,-1}$$

$$\delta' = (\beta_1'\ ,\ \gamma_1')$$

and let $\bar{W}_1$ be a set of instruments

$$\bar{W}_1 = [(\hat{Y}_1 - \hat{r}\ Y_{1,-1})\ ,\ (X_1 - \hat{r}\ X_{1,-1})]$$

where $\hat{r}$ is a consistent estimate of r obtained in the first stage and $\hat{Y}_1 = \Phi\hat{\Pi}_1*$, where $\hat{\Pi}_1*$ is the relevant sub-matrix of $\hat{\Pi}*$ formed using the consistent estimates of the B, Γ and R from the first stage. Fair's version of Brundy and Jorgenson estimator (FBJ) is

----

[18](cont'd)extended their proof to the case with autocorrelation.

[19] Note that so far we have been assuming that the value of the autocorrelation coefficient is known. Only in this section, following Fair, we assume it is unknown and use a consistent estimate of it.

$$\hat{\delta} = (\bar{W}_1'\bar{Z}_1)^{-1}\bar{W}_1'\bar{y}_1 \tag{18}$$

This estimator is consistent and if iterated is fully efficient within the class of limited information estimators.[20]

It must be noted that the FBJ estimator can include lagged endogenous variables. However, in the absence of such variables, it can be modified at the initial stage by generating Y from the ordinary reduced form

$$Y = X\,\Pi + V$$

where

$$\Pi = -\Gamma\,B^{-1}\ ;\ V = U\,B^{-1}$$

Since autocorrelation does not destroy consistency, we can obtain a consistent estimate of Y using $\Pi = -\Gamma B^{-1}$. This simplification can substantially reduce the computations at the initial stage. Its effect on the small sample properties of the FBJ estimator will be investigated later. It should be noted that this modified version of the FBJ (MFBJ) estimator can use all T observations. In this case MFBJ will take the form

$$\dot{Z}_1 = (PY_1\ \ PX_1)\ ,\ \dot{y}_1 = Py_1$$

$$\dot{W}_1 = (P\hat{Y}_1\ \ PX_1)$$

$$\hat{\delta}_1 = (\dot{W}_1'\dot{Z}_1)^{-1}\dot{W}_1'\dot{y}_1 \tag{19}$$

$\hat{Y}_1 = X\hat{\Pi}_1$ where $\hat{\Pi}_1$ is a consistent estimate of Π_1 constructed on the basis of consistent estimates of B and Γ.

----

[20] See Fair(1972), pp.446-47.

B. The Theil Generalized Two Stage Least Squares Approach

A different approach to the estimation of the structural equation with first order autocorrelation was suggested by Theil(1958). In this section we shall discuss the Theil estimator and will consider a number of its variants. One of the important characteristic of the Theil type estimators is that they employ the Prais-Winsten transformation matrix for solving the autocorrelation problem and therefore they utilize all the observations. The asymptotic properties of the estimators using T or (T-1) observations will not be different since the weight of the first observation will vanish as T approaches infinity.[21] However, the employment of the first observation might have some effect on the small sample properties of these estimators.

Theil Estimator

Theil(1958, p.345) suggested a generalized 2SLS (G2SLS) method for estimation of the system represented in (1). His estimator can be summarized as follows:

1- First transform all structural equations using the Prais Winsten transformation matrix P

----

[21] The importance of retaining the first observation is particularly emphasized when the exogenous variables are trended. Effects of trended data on the properties of different estimators was studied by Hannan(1960, pp.114-5, 122-28) and Grennander and Rosenblatt(1957, pp.231-254). However, they were mainly concerned with the asymptotic properties.

$$P\,Y\,B + P\,X\,\Gamma = P\,U = \dot{U}$$

or

$$\dot{Y}\,B + \dot{X}\,\Gamma = \dot{U} \tag{20}$$

The reduced form of the system (20) can be written as

$$\dot{Y} = \dot{X}\,\Pi + \dot{V} \tag{21}$$

2- The next step is the transformation of the structural equation under consideration in a similar fashion

$$\dot{y}_1 = \dot{Y}_1\beta_1 + \dot{X}_1\gamma_1 + \epsilon_1 = \dot{Z}_1\delta_1 + \epsilon_1 \tag{22}$$

3- Finally estimate equation (22) by OLS replacing $\dot{Y}_1$ by its fitted values calculated from (21) as

$$\hat{\delta}_1 = (\hat{\dot{Z}}_1'\hat{\dot{Z}}_1)^{-1}\hat{\dot{Z}}_1'\dot{y}_1$$

where

$$\hat{\dot{Z}}_1 = (\hat{\dot{Y}}_1\ ,\ \dot{X}_1)$$

$$\hat{\dot{Y}}_1 = \dot{X}\,\hat{\Pi}_1 = \dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y}_1$$

To construct Theil's estimator we first need a estimate of the coefficient of autocorrelation to construct the matrix P. A consistent estimate of r can be obtained by applying 2SLS, ignoring the autocorrelation, to the structural equation and calculating r using the Prais-Winsten formula defined in Chapter I.

Theil's estimator is consistent and asymptotically efficient within the class of limited information estimators if and only if the non-zero diagonal elements of the R matrix are all equal, a result recently derived in Buse(1983).

As in the case of regular 2SLS the G2SLS estimator can be interpreted as an instrumental variable estimator. The

instrumental variable version of G2SLS for estimation of the system (1) is as follows:

1- Obtain a consistent estimate of the structural coefficients and thus the autocorrelation parameter.

2- Transform the structural equation under consideration by the Prais-Winsten transformation matrix

$$Py_1 = PY_1\beta_1 + PX_1\gamma_1 + Pu_1$$

or

$$\dot{y}_1 = \dot{Y}_1\beta_1 + \dot{X}_1\gamma_1 + \epsilon = \dot{Z}_1\delta_1 + \epsilon_1$$

3- Let W_1 be a set of instruments such as

$$W_1 = (\hat{\dot{Y}}_1 \; \dot{X}_1)$$

where

$$\hat{\dot{Y}}_1 = \dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y}_1$$

4-The instrumental variable analogue of G2SLS is

$$\hat{\delta}_1 = (W_1'\dot{Z}_1)^{-1}W_1'\dot{y}_1$$

As we shall see later, this estimator is numerically equivalent and has the same asymptotic properties as Theil's G2SLS.

Theil(II) Estimator

Theil's G2SLS first transforms the autocorrelated system to one with no autocorrelation and then treats it as the 2SLS problem.[22] However, another estimator can be constructed on the basis of the same principles but in the

---

[22]For a general approach to the derivation of G2SLS and its relation to the Madansky(1964) and Wickens(1969) estimators see Fisher(1972). It should be noted that we have not considered the latter estimators since they are shown (see Fisher(1972), Wickens(1969)) to be asymptotically less efficient than the Theil G2SLS estimator.

reverse order.[23] This means that one first resolves the simultaneity problem in the system and then tackles the autocorrelation problem. This procedure also results in a consistent estimator, which we shall call Theil(II). The Theil(II) estimator can be defined as follows :

1- First consider the reduced form

$$Y = X\,\Pi + V \tag{23}$$

We can obtain a consistent estimate of the Y's by applying OLS to (23).

2- Second, write the structural equation under consideration as

$$y_1 = \hat{Y}_1\beta_1 + X_1\gamma_1 + u_1 + \hat{V}_1\beta_1 \tag{24}$$

where

$$\hat{Y}_1 = X(X'X)^{-1}X'Y_1$$

3- Finally, transform equation (24) using the matrix P and estimate the resulting equation by OLS

$$\hat{\delta}_1 = (\dot{\hat{Z}}_1{}'\dot{\hat{Z}}_1)^{-1}\dot{\hat{Z}}_1{}'\dot{y}_1$$

where

$$\dot{\hat{Z}}_1 = (P\hat{Y}_1\ ,\ PX_1)$$

As shown in the next chapter, the Theil(II) estimator is consistent and asymptotically equal to the G2SLS estimator.

The instrumental variable version of Theil(II), denoted as Theil(IV), can be constructed as follows:

1- Let $\dot{W}_1$ be a set of instruments such as

----

[23]This estimator and its Instrumental variable analogue were mainly worked out in Buse(1983).

$$\dot{W}_1 = (\dot{\hat{Y}}_1 \ \dot{X}_1)$$

where

$$\hat{Y}_1 = X(X'X)^{-1}X'Y_1 \text{ and } \dot{\hat{Y}}_1 = P\hat{Y}_1$$

2- Transform the structural equation (1) using a consistent estimate of r.

$$Py_1 = PY_1\beta_1 + PX_1\gamma_1 + Pu_1$$

or

$$\dot{y}_1 = \dot{Y}_1\beta_1 + \dot{X}_1\gamma_1 + \epsilon_1 = \dot{Z}_1\delta_1 + \epsilon_1$$

3- Estimate the transformed system by instrumental variables as

$$\hat{\delta}_1 = (\dot{W}_1'\dot{Z}_1)^{-1}\dot{W}_1'\dot{y}_1$$

We shall show later that like Theil(II), Theil(IV) has the same asymptotic properties as Theil's G2SLS estimator.

Modified Generalized Two Stage Least Squares Estimator

A modified Generalized Two Stage Least Squares (MG2SLS) method for estimation of (1) can be constructed as follows:

1- At the first stage, use the ordinary reduced form and regress each column of Y_1 on the set of X's.

2- Using OLS, estimate (1) using the fitted values of Y_1 in place of Y_1. Use the estimated residuals to estimate r and transform the structural equation such as

$$y_1 - \hat{r}\, y_{1,-1} = (Y_1 - \hat{r}\, Y_{1,-1})\beta_1 + (X_1 - \hat{r}\, X_{1,-1})\gamma_1 + \epsilon_1 \quad (25)$$

3- Estimate equation (25) iteratively by OLS replacing $(Y_1 - \hat{r}\, Y_{1,-1})$ by its fitted value from a regression of $(Y_1 - \hat{r}\, Y_{1,-1})$ on exogenous variables transformed as $(X - \hat{r}\, X_{-1})$.

This method, which was suggested by Pindyck and Rubinfeld (1981) is just the Theil estimator which uses the Q instead of the prais winsten transformation matrix. It is consistent and asymptotically equivalent to the Theil's G2SLS.

Generalized Limited Information Maximum Likelihood Estimator

As shown in Appendix I, Sargan and Amemiya have derived (6) under restrictive assumptions which can affect the asymptotic and small sample properties of their estimator. It is also shown in Appendix I that under different set of assumptions one can derive a LIML estimator that has the same asymptotic distribution as Theil's G2SLS. The intuitive explanation of this estimator that we shall call Generalized limited information maximum likelihood (GLIML) can be given in terms of the so-called "least variance ratio principle". For a given value of the autocorrelation coefficient, the structural equation

$$y_1 = Y_1\beta_1 + X_1\gamma_1 + u_1$$

can be transformed, using the Q (Cochrane-Orcutt) matrix, to the equation free of autocorrelation; namely,

$$\bar{y}_1 = \bar{Y}_1\beta_1 + \bar{X}_1\gamma_1 + \epsilon_1 \tag{26}$$

where $\bar{y}_1$, $\bar{Y}_1$ and $\bar{X}_1$ are transformed variables.

Equation (26) can be written as

$$\bar{y}_1 - \bar{Y}_1\beta_1 = (\bar{y}_1 \quad \bar{Y}_1)\begin{pmatrix} 1 \\ -\beta_1 \end{pmatrix} = Y_1{*}\beta_1{*}$$

$$= \bar{y}_1{*} = \dot{X}_1\gamma_1 + \epsilon_1 \tag{27}$$

where

$\beta_1{*}' = (1,-\beta_1')$ and $Y_1{*} = (\bar{y}_1\ \bar{Y}_1)$

The "composite dependent variable" $\bar{y}_1{*}$ is a linear combination of the transformed endogenous variables included in the first structural equation. For given β_1, the composite variable $\bar{y}_1{*}$ can be calculated and the coefficient γ_1 can be estimated as

$$\hat{\gamma}_1 = (\bar{X}_1'\bar{X}_1)^{-1}\bar{X}_1'\bar{y}_1{*} \tag{28}$$

Therefore, the sum of squared residuals of the equation (27) will be

$$\zeta_1 = \hat{v}_1{*}'\ \hat{v}_1{*}$$

where

$$v_1{*} = \bar{y}_1{*} - \bar{y}_1{*} = (I-\bar{X}_1(\bar{X}_1'\bar{X}_1)^{-1}\bar{X}_1')\bar{y}_1{*}$$

therefore

$$\zeta_1 = \bar{y}_1{*}'(I-\bar{X}_1(\bar{X}_1'\bar{X}_1)^{-1}\bar{X}_1')\bar{y}_1{*}$$

If the set of predetermined variables in the structural equation (27) is extended to include all the predetermined variables of the system, we shall have

$$\bar{y}_1{*} = \bar{X}\ \gamma + \epsilon_1 \tag{29}$$

where $\bar{X} = (\bar{X}_1\ \bar{X}^{x})$ and $\gamma' = (\gamma_1'\ 0)$ and $\bar{X}^{x}$ is the set of transformed predetermined variables excluded from the first structural equation. If we ignore the knowledge about the structure of the coefficients γ and apply OLS to equation (29) we will have

$$\hat{\gamma} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{y}_1{*}$$

The sum of squared residuals corresponding to the regression

of the composite variable $\overline{y}_1*$ on all the transformed predetermined variables $\overline{X}$ will be equal to

$$\zeta_2 = \overline{y}_1*(I-\overline{X}(\overline{X}'\overline{X})^{-1}\overline{X})\overline{y}_1*$$

The estimate of the coefficient β_1 can be obtained by finding the values of β_1 which minimizes the effect of the additional transformed predetermined variables $\overline{X}^x$ on the sum of squared error term. In other words, the estimated β_1 are the values which minimizes the variance ratio

$$L = \zeta_1 / \zeta_2$$

L is always greater or equal to one, since the additional predetermined variables can only decrease the sum of squared errors term.

L can be written as

$$L = (\beta_1*'\pi_1\beta_1*)/ (\beta_1*'\pi_2\beta_1*) \qquad (30)$$

where

$$\pi_1 = Y_1*'(I-\overline{X}_1(\overline{X}_1'\overline{X}_1)^{-1}\overline{X}_1')Y_1*$$

$$\pi_2 = Y_1*'(I-\overline{X}(\overline{X}'\overline{X})^{-1}\overline{X}')Y_1*$$

where β_1* and Y_1* are defined above.

In Appendix I we show that minimization of L with respect to β_1 and then estimation of γ_1 from (28) will give the same numerical and analytical results as those obtained by partial differentiation of the following concentrated likelihood function with respect to β_1 and γ_1 and solving the resultant first order conditions,

$$L_2 = K* + 1/2 \log(\beta_1*'\pi_2\beta_1*)-1/2 \log|\pi_2|-1/2 \log(\epsilon'\epsilon) \qquad (31)$$

Derivation of the concentrated likelihood function (31) and its difference with the one derived by Sargan and

Amemiya is also given in the Appendix I. Moreover, it is shown in Chapter III that under the assumption of equal autocorrelation coefficients across equations the ALIML and GLIML estimators are asymptotically equivalent.

To carry out the GLIML procedure, we need to know the value of the autocorrelation coefficient r. A consistent estimate of r can be obtained by partial differentiation of the concentrated likelihood function (31) with respect to r and solving the resultant first order condition for r. However, this procedure resulted in a very complicated formula which was not of any practical value.[24] As an alternative, a search over r on the interval (-1,+1) was made and estimates of r, β_1 and γ_1 which minimize the sum of squared residuals were chosen. It is clear that the GLIML estimator can be extended to employ all T observations. The interpretation of this version in terms of the so-called least variance ratio principle is the same as what was presented for the case with T-1 observations. Derivation of the concentrated likelihood function in this case is the same as the one presented in Appendix I except that in this case the Prais-Winsten transformation matrix will be employed for transforming the variables. The log-likelihood

----

[24] The second alternative was to calculate r from one of the three procedures: CORC, Prais-Winsten or the Cubic formula developed by Beach and Mackinnon. Pilot experiments showed that these alternatives often made the concentrated log-likelihood function reach one of its local maxima instead of the global one. This result can be explained by the fact that none of the above three methods are the direct solution to the first order condition but they are only approximations.

function of the system in this case will have an extra term which is the Jacobian of the transformation. This Jacobian does not cancel out in the process of concentrating the log-likelihood function to incorporate the coefficients of the first structural equation. The concentrated log-likelihood function in this case is

$$L_2=K^* +|J|+1/2 \log(\beta_1{}^{*\prime}\pi_2 \ \beta_1{}^*)-1/2 \log|\pi_2|-1/2 \log(\epsilon'\epsilon) \quad (32)$$

which is the extension of the formula given in (31) except for an extra term, i.e., the Jacobian of the transformation, and the fact that in (32) variables were transformed by the P matrix instead of the Cochrane-Orcutt transformation used in (31).

Other Estimators

Other limited information estimators have been proposed for estimation of autoregressive simultaneous equation system. Two of these methods that we shall review here are those proposed by Klein(1974) and Kmenta(1971).

1. **The Klein Estimator**

An alternative approach for estimation of the system presented in (1) and (2) was suggested by Klein(1974, p.207-210). He argued against the employment of the augmented reduced form on the grounds that the enlargement of the vector of reduced form regressors taxes the degrees of freedom and employment of X and X_{-1} as regressors raises the possibility of strong multicollinearity. He suggested

using the ordinary reduced form, while taking into account the information that the reduced form disturbances satisfy an autoregressive scheme. His proposed method amounts to an iterative scheme using the generalized least squares method at the initial stage. Then, using the fitted values of the jointly dependent variables from the first stage, estimate iteratively the transformed equation (4). The main problem with this approach is in the estimation of the variance-covariance matrix of the reduced form at the first stage. Write the ordinary reduced form (equation (11), chapter I) as

$$Y = X \Pi + V \tag{33}$$

where

$$V = U B^{-1}$$

and

$$V_t = (u_{t1}, \ldots, u_t \;) B^{-1}$$

or

$$V_t = (\sum_{j=1}^{G} b^{j1} u_{tj}, \ldots, \sum_{j=1}^{G} b^{jG} u_{tj})$$

Where b^{ij} is the (i,j) element of the inverse of B.

The ith reduced form disturbance is equal to

$$V_{ti} = \sum_{j=1}^{G} b^{ji} u_{tj} = \sum_{j=1}^{G} b^{ji} \; r_j u_{t-1,j} + \sum_{j=1}^{G} b^{ji} \epsilon_{tj}$$

or

$$V_{ti} = \sum_{j=1}^{G} b^{ji} \epsilon_{tj} + \ldots + \sum_{j=1}^{G} b^{ji} \, r_j^{s} \; \epsilon_{t-s,j} + \ldots$$

or

$$V_{ti} = \sum_{s=0}^{\infty} \sum_{j=1}^{G} b^{ji} \; r_j^{s} \; \epsilon_{t-s,j}$$

Therefore

$$E(V_{ti}) = 0$$

$$E(V_{ti}^2) = \sigma^2 \{\sum_{j=1}^{G}[(b^{ji\,2})/(1-r_j^2)]\} \tag{34}$$

so that equation (34) shows that the estimation of the variance of the ith reduced form disturbance requires the knowledge of the autocorrelation coefficients in all the equations of the system. Only under the restrictive assumption that the diagonal elements of R matrix are all equal can the variance-covariance matrix of the reduced form disturbances be easily calculated.

Klein, however, overlooked this problem and proposed the following procedure for estimation of the variance-covariance matrix of the reduced form disturbances:

1- Regress Y_t on the set of exogenous variables.

2- Use the residuals of the first step to compute the covariance matrix as

$$\hat{\Omega} = \begin{pmatrix} \hat{V}_{1i} \\ \cdot \\ \\ \cdot \\ \hat{V}_{Ti} \end{pmatrix} (\hat{V}_{1i},\dots,\hat{V}_{Ti}) = \begin{pmatrix} \hat{V}_{1i}^2,\dots,\hat{V}_{1i}\hat{V}_{Ti} \\ \\ \\ \hat{V}_{Ti}\ V_{1i},\dots,\hat{V}\ _i^2 \end{pmatrix}$$

A feasible GLS estimator requires $\hat{\Omega}_i^{-1}$ but it is clear that matrix $\hat{\Omega}_i$ is singular with rank one and therefore the Klein method is not operational.

2. The Kmenta Estimator

Kmenta(1971, pp.587-589) also proposed a method for estimation of the model (1). He explicitly assumed the

absence of lagged endogenous variables in his model. His estimator is as follows:

1- Using the augmented reduced form, regress each column of Y_1 on the lagged endogenous variables as well as current and lagged values of the exogenous variables using OLS.

2- Using the fitted values of the Y_1 and $Y_{1,-1}$ in the transformed equation (3) as

$$y_1 - r\, y_{1,-1} = (\hat{Y}_1 - r\, \hat{Y}_{1,-1})\beta_1 + (X_1 - r\, X_{1,-1})\gamma_1 + \epsilon_1 \quad (35)$$

Estimating the coefficients r, β_1, γ_1 by applying the restricted least squares method to (35).

The list of instruments in Kmenta's method is the same as Sargan's and thus is subject to the same criticisms; especially if the X's are subject to some sort of trend, the appearance of X's and X_{-1}'s among the instruments might create severe multicollinearity. Moreover, obtaining the fitted values of the endogenous variables from the augmented reduced form automatically omits one observation. Therefore, his estimator loses another observation due to the employment of $\hat{Y}_{1,-1}$ in the second stage regression. In empirical work, where researchers usually have a limited number of observations losing two observations is not desirable.

Table2-1 presents the general structure of the estimators discussed in this chapter. Table2-2 shows the mathematical presentation of Table2-1. The computational summary of each of the estimators is given in Appendix II.

Table 2-1: Simultaneity and Autocorrelation

$y = Y B + X C u \; ; \; u = r u + e$

Ordinary reduced form (O.R.F) (T or T-1 observations)		Augmented Reduced Form (A.R.F) (T-1 Observations)	O.R.F or A.R.F (T or T-1 Observations)		
Prais Winsten (T)	CORC (T-1)	CORC (T-1)	P.W (T)	CORC (T-1)	CORC (T-1)
Theil(G2SLS)	Rubinfeld and Pindyk (MG2SLS)	Fair*	GLIML (P.W)	GLIML (CORC)	ALIML
Theil(IV)		Kmenta			
Theil(II)		Fair Brundy			
Modified Brundy& Jorgenson					

*Note that the Fair estimator only uses a sub-set of the variables appearing in the augmented reduced form.

Table 2-2: Simultaneity and Autocorrelation
(Mathematical Demonstration)

Ordinary Reduced Form(O.R.F)			Augmented Reduced Form(A.R.F)	O.R.F (LIML)	A.R.F (LIML)
Prais-Winsten Formula		Cochrane-Orcutt	Cochrane-Orcutt		
G L S	O L S	G L S			
$\hat{\Pi}_1 = (\dot{X}'\dot{X})^{-1} \dot{X}'\dot{Y}_1$	$\hat{\Pi}_1 = (X'X)^{-1} X'Y_1$	$\hat{\Pi}_1 = (\dot{X}'\dot{X})^{-1} \dot{X}'\dot{Y}_1$	$\hat{\Pi}_1 = (\hat{\Gamma}\hat{B}\hat{R})_1$	GLIML	ALIML (1966)
$\widehat{P Y_1} = \dot{X}\hat{\Pi}_1$	$P\hat{Y}_1 = P X \hat{\Pi}_1$	$\widehat{P Y_1} = \dot{X}\hat{\Pi}_1$	Fair Brundy and Jorgenson (1972)		
Theil(G2SLS) (1961)	Theil(IV)	Rubinfeld & Pindyck(MG2SLS) (1979)			
	Theil(II)		Kmenta(1971)		
			Fair(1970)		

O.R.F : $Y = X \Pi + V$

A.R.F : $Y = X \Pi_1 + Y_{-1} \Pi_2 + X_{-1} \Pi_3 + E$

G L S : Generalized Least Squares

O L S : Ordinary Least Squares

III. ASYMPTOTIC PROPERTIES OF THE ESTIMATORS

This chapter compares the asymptotic properties of different estimators discussed earlier. We will assume that the autocorrelation coefficients are known to prevent complications which can arise due to the substitution of the consistent estimates for the unknown autocorrelation coefficients. This is in fact a common practice.[25] It is known from general statistical principles that if autocorrelation coefficient is unknown a greater uncertainty (variability) will be introduced for all the coefficients and we can conjecture that this effect will be roughly the same across all estimators.

The order of the discussion of the estimators will be the same as in Chapter 2. We shall first present the asymptotic properties of the ALIML estimator and then use that as a base for comparison.

The ALIML Estimator

Consider the following simultaneous equation system

$$Y B + X \Gamma = U \tag{1}$$

$$U = U_{-1} R + E$$

$$E(E_t' E_t) = \Theta$$

Let the structural equation under consideration be

$$y_1 = Y_1 \beta_1 + X_1 \gamma_1 + u_1$$

or

---

[25] See for example Wickens(1969), Fair(1970,1972) and Fisher(1972).

$$y_1 = Z_1\delta_1 + u_1 \tag{2}$$

$$u_1 = r_1 \, u_{1,-1} + \epsilon_1$$

$$E(\epsilon_1\epsilon_1') = \sigma_{11} \, I$$

$$E(u_1u_1') = \sigma_{11} \, \Sigma$$

where

$Z_1 = (Y_1, X_1)$, and $\delta_1' = (\beta_1' \quad \gamma')$

In the subsequent discussion we will make the usual assumptions that Plim $(X'E/T) = 0$, and Plim $(X'\Sigma^{-1}X/T)$ is a K.K positive definite matrix.

Denote

$$Y_1* = (y_1 \quad Y_1)$$

The ALIML estimator is equal to

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{pmatrix} = \begin{pmatrix} \overline{Y}_1'\overline{Y}_1 - \lambda\overline{Y}_1'(I-\Phi(\Phi'\Phi)^{-1}\Phi')\overline{Y}_1 & \overline{Y}_1'\overline{X}_1 \\ \overline{X}_1'\overline{Y}_1 & \overline{X}_1'\overline{X}_1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \overline{Y}_1 - \lambda\overline{Y}_1(I-\Phi(\Phi'\Phi)^{-1}\Phi') \\ \overline{X}_1' \end{pmatrix} \overline{y}_1$$

where λ is the smallest root of the determinantal equation

$$|\pi_1* - \lambda \; \pi*| = 0$$

and

$$\pi_1* = \overline{Y}_1*'(I-\overline{X}_1(\overline{X}_1'\overline{X}_1)^{-1}\overline{X}_1')\overline{Y}_1*$$

$$\pi* = Y_1*'(I-\Phi(\Phi'\Phi)^{-1}\Phi')Y_1*$$

and

$$\Phi = (X \; , \; Y_{-1} \; , \; X_{-1})$$

Amemiya showed that ALIML is consistent and its asymptotic covariance matrix is equal to

$$\text{Asy-cov}\ \sqrt{T}\begin{pmatrix}\hat{\beta}_1-\beta_1\\ \\ \hat{\gamma}_1-\gamma_1\end{pmatrix} =$$

$$\sigma_{11}\text{plim}\begin{pmatrix}\bar{Y}_1\Phi(\Phi'\Phi)^{-1}\Phi'\bar{Y}_1 & \bar{Y}_1'\bar{X}_1\\ \\ \bar{X}_1'\bar{Y}_1 & \bar{X}_1'\bar{X}_1\end{pmatrix}^{-1} \quad (3)$$

Fair's Estimator

Fair has not derived the asymptotic distribution of his estimator.[26] We do not do so either but we do provide an alternative proof of consistency to that given by Fair.

To prove the consistency of Fair's estimator, for known r, we can combine equations (10) and (14) of Chapter 2 to obtain

$$\begin{pmatrix}\hat{\beta}_1\\ \hat{\gamma}_1\end{pmatrix} = \begin{pmatrix}\bar{Y}_1'X_m(X_m'X_m)^{-1}X_m'\bar{Y}_1 & \bar{Y}_1'\bar{X}_1\\ \bar{X}_1'\bar{Y}_1 & \bar{X}_1'\bar{X}_1\end{pmatrix}^{-1} \cdot \begin{pmatrix}\bar{Y}_1'X_m(X_m'X_m)^{-1}X_m'\\ \bar{X}_1'\end{pmatrix}\bar{y}_1$$

Substituting for $\bar{y}_1$, we will have

$$\begin{pmatrix}\hat{\beta}_1\\ \hat{\gamma}_1\end{pmatrix} = \begin{pmatrix}\beta_1\\ \gamma_1\end{pmatrix} + \begin{pmatrix}\bar{Y}_1'X_m(X_m'X_m)^{-1}X_m'\bar{Y}_1 & \bar{Y}_1'\bar{X}_1\\ \bar{X}_1'\bar{Y}_1 & \bar{X}_1'\bar{X}_1\end{pmatrix}^{-1} \cdot \begin{pmatrix}\bar{Y}_1'X_m(X_m'X_m)^{-1}X_m'\epsilon_1\\ \bar{X}_1'\epsilon_1\end{pmatrix}$$

---

[26] Actually Fair proposed a number of different alternative estimators in his paper. In this thesis we take the Fair estimator to mean the three step estimator outlined in Chapter 2.

It is easy to show that all the plims involving $\overline{Y}_1'X_m$ exist. Furthermore, we know by assumption that

$$\text{Plim}(\overline{X}_1'\epsilon_1/ T) = 0$$

$$\text{Plim}(X_m'\epsilon_1/ T) = 0$$

Hence we have

$$\text{Plim}\begin{pmatrix}\hat{\beta}_1 \\ \hat{\gamma}_1\end{pmatrix} = \begin{pmatrix}\beta_1 \\ \gamma_1\end{pmatrix}$$

which proves the consistency of the Fair estimator.

We showed above that the Fair estimator is identical to S2SLS estimator if it uses all the instruments appearing in the augmented reduced form of the system under consideration. However, it uses only a sub-set of those instruments and therefore is asymptotically less efficient than the ALIML and S2SLS estimators.

6- *The Brundy and Jorgenson Estimators*

The Fair version of Brundy and Jorgenson's estimator utilizes the augmented reduced form where the Π's are formed using the consistent estimates of B, Γ, and R. Fair(1972) showed that this estimator is consistent and asymptotically efficient within the class of limited information estimators.

To utilize all observations, we introduced a modified version of the Brundy and Jorgenson estimator which uses the ordinary reduced form. The proof of the consistency of this estimator is the same as for the instrumental variable version of the Theil(II) estimator considered below.

Following the same procedure we can obtain the asymptotic distribution of the modified Brundy and Jorgenson estimator as

$$\sqrt{T}\,(\hat{\delta}_1-\delta_1) \sim N[0,\sigma_{11}\text{plim}(W_3'\dot{Z}_1)^{-1}]$$

where

$$W_3 = [\,P\hat{Y}_1\,,\;P\,X_1]$$
$$\dot{Z}_1 = (\dot{Y}_1\;\dot{X}_1)$$

and $\hat{Y}_1$ is the fitted values of Y_1 calculated from the consistent estimate of the ordinary reduced form.

Since $\hat{Y}_1$ is calculated from a consistent estimate of the reduced form coefficients, the asymptotic variance covariance matrix of the Brundy and Jorgenson estimator when it uses the ordinary reduced form would be equal to Theil's G2SLS that is discussed later in this chapter.[27]

Comparison of the modified Brundy and Jorgenson estimator when it uses T or T-1 observations will demonstrate the small sample efficiency gain due to the utilization of the first observation. On the other hand comparing it with Fair's Brundy and Jorgenson estimator which also uses T-1 Observations will show the small sample differences due to employment of two different reduced forms. Finally, comparison of the modified Brundy and Jorgenson estimator when it uses all T observations with the Fair version shows the small sample efficiency differences of using the augmented reduced form while losing one

---

[27] As we shall see the Theil estimator is generally less efficient than the ALIML estimator. Therefore, the BJ estimator will, in general, be less efficient than the ALIML estimator.

observation, against using the ordinary reduced form while retaining the first observation.

Theil's Estimator (G2SLS)

In this and the subsequent sections we shall deal with the asymptotic properties of the Theil estimator and its variants. We will go into a detailed discussion of the properties of these estimators simply because the literature is incomplete on this topic.

The reduced form of the system (1) can be written as

$$Y = X\,\Pi + V \tag{4}$$

Transform (4) by the Prais-Winsten matrix P

$$P\,Y = P\,X\,\Pi + P\,V$$

or

$$\dot{Y} = \dot{X}\,\Pi + \dot{V} \tag{5}$$

Apply OLS to the system (5) to obtain the fitted values of Y_1 as

$$\hat{\dot{Y}}_1 = \dot{X}\,\hat{\Pi}_1 = \dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y}_1$$

Then Theil's G2SLS estimator of δ_1 is

$$\hat{\delta}_1 = \begin{pmatrix} \dot{Y}_1'\dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y}_1 & \dot{Y}_1'\dot{X}_1 \\ \dot{X}_1'\dot{Y}_1 & \dot{X}_1'\dot{X}_1 \end{pmatrix}^{-1} \begin{pmatrix} \dot{Y}_1'\dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}' \\ \dot{X}_1' \end{pmatrix} \dot{y}_1 \tag{6}$$

The consistency of the Theil estimator for the case with known and unknown autocorrelation coefficient is demonstrated by Wickens. The asymptotic covariance matrix of the G2SLS estimator when r is known is given by

$$\text{Asy-cov}\ \sqrt{T}\begin{pmatrix}\hat{\beta}_1-\beta_1\\ \\ \hat{\gamma}_1-\gamma_1\end{pmatrix} =$$

$$\sigma_{11}\operatorname{plim}\begin{pmatrix}\dot{Y}_1\dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y}_1 & \dot{Y}_1'\dot{X}_1\\ \\ \dot{X}_1'\dot{Y}_1 & \dot{X}_1'\dot{X}_1\end{pmatrix}^{-1} \tag{7}$$

$$= \ \operatorname{plim}\begin{pmatrix}(\Pi_1'\dot{X}'\dot{X}\Pi_1/T) & (\Pi_1'\dot{X}'\dot{X}_1/T)\\ \\ (\dot{X}_1'\dot{X}\ \Pi_1/T) & (\dot{X}_1'\dot{X}_1/T)\end{pmatrix} = \begin{pmatrix}\Pi_1'\dot{\Psi}\ \Pi_1 & \Pi_1'\dot{\Psi}_1\\ \\ \dot{\Psi}_1'\Pi_1 & \dot{\Psi}_{11}\end{pmatrix} = \Delta \tag{8}$$

where

$$\dot{\Psi} = \operatorname{plim}(X'X/T) \tag{9}$$

$$\dot{\Psi}_1 = \operatorname{plim}(X'X_1/T) \tag{10}$$

$$\dot{\Psi}_{11} = \operatorname{plim}(X_1'X_1/T) \tag{11}$$

are assumed to exist.

It is clear that the asymptotic covariance matrix of the Theil estimator given by equation (8) is numerically different from the asymptotic covariance matrix of ALIML estimator (equation (3)). The numerical difference of ALIML and Theil's G2SLS covariance matrices raises the question of which one is efficient.

At this point we should note that G2SLS is constructed using T observations while ALIML utilizes only T-1 observations. However, since the weight of the first observation is asymptotically negligible, in the following discussion we shall assume that all variables are transformed by the Q (Cochrane-Orcutt) matrix.

The relative efficiency of ALIML and G2SLS has been investigated in Buse(1983), which we follow here. First introduce a selection matrix S_x such that

$$\Phi S_x = X$$

where $S_x = (I \ , \ 0 \ , \ 0)$

Therefore, we can write $\overline{X}$ as

$$\overline{X} = X - r_1 X_{-1} = \Phi(S_x - r_1 S_{x,-1}) \tag{12}$$

Now using the reduced form $\overline{Y}_1 = \overline{X}\,\Pi_1 + \overline{V}_1$, the first diagonal term in the asymptotic covariance matrix (3) will become

$$\overline{Y}_1'\Phi(\Phi'\Phi)^{-1}\Phi'\overline{Y}_1 = (\overline{X}\,\Pi_1 + \overline{V}_1)'\Phi(\Phi'\Phi)^{-1}\Phi'(\overline{X}\,\Pi_1 + \overline{V}_1) \tag{13}$$

Using the result in (12), equation (13) will become

$$\begin{aligned}\overline{Y}_1'\Phi(\Phi'\Phi)^{-1}\Phi'\overline{Y}_1 &= \Pi_1'\overline{X}'\overline{X}\,\Pi_1 + \Pi_1'\overline{X}'\overline{V}_1 \\ &\quad + \overline{V}_1'\overline{X}\,\Pi_1 + \overline{V}_1'\Phi(\Phi'\Phi)^{-1}\Phi'\overline{V}_1\end{aligned} \tag{14}$$

Taking probability limit of each term in (14), we shall have

$$\text{Plim}\ (\Pi_1'\overline{X}'\overline{X}\,\Pi_1/T) = \Pi_1'\overline{\Psi}\,\Pi_1 \tag{15}$$

$$\text{Plim}\ (\Pi_1'\overline{X}'\ \overline{V}_1/T) = 0 \tag{16}$$

$$\text{Plim}\ (\overline{V}_1'\Phi(\Phi'\Phi)^{-1}\Phi'\overline{V}_1/T) = \Upsilon'(\Phi'\Phi)^{-1}\Upsilon \tag{17}$$

where

$$\Upsilon = \text{plim}\ (\Phi'\ \overline{V}_1/T) \neq 0$$

since

$$\begin{aligned}\Phi'\overline{V}_1 &= (X - r_1X_{-1}\ ,\ Y_{-1} - r_1Y_{-2}\ ,\ X_{-1} - r_1X_{-2})' \\ &\quad .(U - r_1U_{-1})(B^{-1})_1'\end{aligned}$$

and therefore Y_{-1} and Y_{-2} will be correlated with U and U_{-1} and thus Υ will be non-zero. Using the results in (15), (16) and (17) and taking the probability limit of the off-diagonal terms of the equation (3), the asymptotic

covariance of ALIML will become

$$\text{Asy-Cov(ALIML)} = \psi = \sigma_{11} \begin{pmatrix} \Pi_1{}'\bar{\Psi}\,\Pi_1 + \Upsilon'(\Phi'\Phi)\,\Upsilon & \Pi_1{}'\bar{\Psi}_1 \\ \bar{\Psi}_1{}'\,\Pi_1 & \bar{\Psi}_{11} \end{pmatrix}^{-1} \quad (18)$$

where $\bar{\Psi}$, $\bar{\Psi}_1$ and $\bar{\Psi}_{11}$ are defined before.

G2SLS is efficient relative to ALIML only if $\Delta-\psi$ is positive semi-definite matrix. This will be so if and only if $\psi^{-1}-\Delta^{-1}$ is positive semi-definite. Comparing (18) with the asymptotic covariance matrix of G2SLS, Δ, defined in (8) we shall have

$$\psi^{-1} - \Delta^{-1} = \begin{pmatrix} \Upsilon'(\Phi'\Phi)\,\Upsilon & 0 \\ 0 & 0 \end{pmatrix}$$

which is a positive semi-definite matrix. Therefore ALIML is at least as efficient as Theil's G2SLS estimator. Only under the condition that all the autocorrelation coefficients are equal we shall have

$$U - r_1 U_{-1} = E$$

and therefore by the assumption that E is uncorrelated with Y_{-1} and Y_{-2} and thus the plim($\Phi'\,\bar{V}_1/T$) will be equal to zero that means the G2SLS and ALIML estimators are identically efficient.[28]

----

[28] Wickens(1969) has argued that the Theil estimator with known or estimated autocorrelation coefficient is efficient within the class of limited information. However, the above proof shows the fallacy of his argument.

An Instrumental Variable Interpretation of the G2SLS Estimator

As was mentioned above, Theil's G2SLS estimator can be interpreted as an instrumental variable estimator. Demonstration of this property, even though straightforward, can shed some light on the general approach of the Theil type estimators and therefore can be useful for interpretations.

Consider estimation of the first structural equation. Let $\dot{W}_1$ be a set of instruments such as

$$\dot{W}_1 = (\hat{\dot{Y}}_1 \quad \dot{X}_1)$$

where

$$\hat{\dot{Y}}_1 = \dot{X} (\dot{X}' \dot{X})^{-1} \dot{X}' \dot{Y}_1$$

and

$$\dot{Y}_1 = P Y_1 = P X \Pi_1 + P V_1 = \dot{X} \Pi + \dot{V}_1$$

Note that

$$\dot{Y}_1 = \hat{\dot{Y}}_1 + \hat{\dot{V}}_1$$

and from the properties of OLS we have the following orthogonality conditions

$$\hat{\dot{Y}}_1{}' \hat{\dot{V}}_1 = 0$$
$$\dot{X}_1{}' \hat{\dot{V}}_1 = 0$$

The structural equation under consideration can be written as

$$P y_1 = P Y_1 \beta_1 + P X_1 \gamma_1 + P u_1$$

or

$$\dot{y}_1 = \dot{Z}_1 \delta_1 + \dot{u}_1$$

Therefore, the instrumental variable version of G2SLS will

be

$$\hat{\delta}_1 = (\dot{W}_1'\dot{Z}_1)^{-1}\dot{W}_1'\dot{y}_1 \tag{19}$$

To prove the equivalence of this estimator and the G2SLS estimator we first note that $\dot{W}_1'\dot{Z}_1$ can be written as

$$\dot{W}_1'\dot{Z}_1 = \begin{pmatrix} \dot{Y}_1' \\ \\ \dot{X}_1' \end{pmatrix} (\dot{Y}_1 \ \ \dot{X}_1) = \begin{pmatrix} \hat{\dot{Y}}_1'\dot{Y}_1 & \hat{\dot{Y}}_1'\dot{X}_1 \\ \\ \dot{X}_1'\dot{Y}_1 & \dot{X}_1'\dot{X}_1 \end{pmatrix}$$

$$= \begin{pmatrix} \hat{\dot{Y}}_1'(\hat{\dot{Y}}_1+\hat{\dot{V}}_1) & \hat{\dot{Y}}_1'\dot{X}_1 \\ \\ \dot{X}_1'(\hat{\dot{Y}}_1+\hat{\dot{V}}_1) & \dot{X}_1'\dot{X}_1 \end{pmatrix} = \begin{pmatrix} \hat{\dot{Y}}_1'\hat{\dot{Y}}_1 & \hat{\dot{Y}}_1'\dot{X}_1 \\ \\ \dot{X}_1'\hat{\dot{Y}}_1 & \dot{X}_1'\dot{X}_1 \end{pmatrix} \tag{20}$$

since $\hat{\dot{Y}}_1'\hat{\dot{V}}_1$ and $\dot{X}_1'\hat{\dot{V}}_1$ are equal to zero given the orthogonality properties of OLS. The term $\dot{W}_1'\dot{y}_1$ can also be expanded as

$$\dot{W}_1'\dot{y}_1 = \begin{pmatrix} \hat{\dot{Y}}_1' \\ \\ \dot{X}_1' \end{pmatrix} \dot{y}_1 = \begin{pmatrix} \hat{\dot{Y}}_1'\dot{y}_1 \\ \\ \dot{X}_1'\dot{y}_1 \end{pmatrix} \tag{21}$$

Using (20) and (21), reduces (19) to the original Theil estimator

$$\hat{\delta}_1 = \begin{pmatrix} \hat{\dot{Y}}_1'\hat{\dot{Y}}_1 & \hat{\dot{Y}}_1'\dot{X}_1 \\ \\ \dot{X}_1'\hat{\dot{Y}}_1 & \dot{X}_1'\dot{X}_1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\dot{Y}}_1'\dot{y}_1 \\ \\ \dot{X}_1'\dot{y}_1 \end{pmatrix}$$

An Alternative Instrumental Variable G2SLS Estimator(Theil(IV))

The instrumental variable interpretation of Theil's estimator uses the estimated transformed endogenous variables as instruments. However, we can introduce another instrumental variable estimator which uses the transformed estimated endogenous variables as instruments. It turns out that this estimator is asymptotically equivalent to Theil's G2SLS estimator. We will call this estimator Theil(IV).

Consider again the first structural equation under consideration

$$y_1 = Y_1\beta_1 + X_1\gamma_1 + u_1 = Z_1\delta_1 + u_1 \tag{22}$$

where Z_1 and δ_1 are defined as before.

Transform equation (22) using the Prais-Winsten transformation matrix to obtain

$$\dot{y}_1 = \dot{Z}_1\delta_1 + \dot{u}_1 \tag{23}$$

To estimate equation (23), denote $\dot{W}_2$ as a set of instruments defined by

$$\dot{W}_2 = (\dot{\hat{Y}}_1 \quad \dot{X}_1)$$

Where

$$\hat{Y}_1 = X(X'X)^{-1}X'Y_1 \tag{24}$$

$$\hat{V}_1 = Y_1 - \hat{Y}_1 = (I - X(X'X)^{-1}X')V_1$$

$$\dot{\hat{Y}}_1 = P\hat{Y}_1 = PX(X'X)^{-1}X'Y_1 = PX\Pi_1 + PX(X'X)^{-1}X'V_1 \tag{25}$$

$$PY_1 = P\hat{Y}_1 + P\hat{V}_1 = \dot{\hat{Y}}_1 + \dot{\hat{V}}_1 \tag{26}$$

The instrumental variable estimator of equation (23) will be equal to

$$\hat{\delta}_{1V} = (\dot{W}_2'\dot{Z}_1)^{-1}\dot{W}_2'\dot{y}_1 \tag{27}$$

which can also be written as

$$\hat{\delta}_{1V} = (\dot{W}_2'\dot{Z}_1)^{-1}\dot{W}_2'(\dot{Z}_1\delta_1 + \dot{u}_1) = \delta_1 + (\dot{W}_2'\dot{Z}_1)^{-1}\dot{W}_2'\dot{u}_1 \tag{28}$$

Therefore, we have

$$\text{Plim}(\hat{\delta}_{1V}) = \delta_1 + \text{Plim}(\dot{W}_2'\dot{Z}_1/T)^{-1}(\dot{W}_2'\dot{u}_1/T) \tag{29}$$

For consistency we need to show that the second term on the right hand side of (29) is equal to zero. To do this, we first expand $\dot{W}_2'\dot{Z}_1$ to yield

$$\text{plim}(\dot{W}_2'\dot{Z}_1/T) = \text{plim}\begin{pmatrix} (\hat{Y}_1'P'P\ Y_1/T) & (\hat{Y}_1'P'P\ X_1/T) \\ (X_1'P'P\ Y_1/T) & (X_1'P'P\ X_1/T) \end{pmatrix} \tag{30}$$

Replacing Y_1 by $Y_1 + V_1$ we will get

$$\text{Plim}\ (\dot{W}_2'\dot{Z}_1/T) =$$

$$\text{Plim}\begin{pmatrix} (\hat{Y}_1'P'P\ \hat{Y}_1/T) + (\hat{Y}_1'P'P\ \hat{V}_1/T) & (\hat{Y}_1'P'P\ X_1/T) \\ (X_1'P'P\ \hat{Y}_1/T) + (X_1'P'P\ \hat{V}_1/T) & (X_1'P'P\ X_1/T) \end{pmatrix} = \Delta^{-1} \tag{31}$$

since following Wickens(1969) we have

$$\text{plim}(\hat{Y}_1'P'P\ \hat{V}_1/T) = \text{plim}(X_1'P'P\ \hat{V}_1/T) = 0$$

and Δ is the covariance matrix of the original Theil G2SLS estimator.

Now consider $\text{plim}(\dot{W}_2'u_1/T)$. We can write this term as

$$\text{plim}(\dot{W}_2'\dot{u}_1/T) = \text{plim}\begin{pmatrix} \dot{\hat{Y}}_1'\dot{u}_1/T \\ \dot{X}_1'\dot{u}_1/T \end{pmatrix} = \begin{pmatrix} \text{plim}\hat{\Pi}_1'\dot{X}'\dot{u}_1/T \\ \text{plim}\ \dot{X}_1'\dot{u}_1/T \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{32}$$

Substituting (31) and (32) in (29) proves the consistency of this estimator.

To investigate the asymptotic distribution of this estimator, using equation (28), we have

$$\sqrt{T}\,(\hat{\delta}_{1V}-\delta_1) = (\dot{W}_2'\dot{Z}_1/T)^{-1}\,\dot{W}_2'\dot{u}_1/\sqrt{T} \qquad (33)$$

Therefore, the asymptotic distribution of $\sqrt{T}(\hat{\delta}_{1V} - \delta_1)$ will be equal to the $\operatorname{plim}(\dot{W}_2'\dot{Z}_1/T)^{-1}$ times asymptotic distribution of $(\dot{W}_2'\dot{u}_1/\sqrt{T})$. Following Theil(1971,pp.380-381) we can show that

$$(\dot{W}_2'\dot{u}_1/\sqrt{T}) \sim N(0,\ \sigma_{11}\operatorname{plim}(\dot{W}_2'\dot{W}_2/T))$$

The probability limit of $(\dot{W}_2'\dot{W}_2/T)$ can be written as

$$\operatorname{plim}(\dot{W}_2'\dot{W}_2/T) = \operatorname{plim}\begin{pmatrix} (\hat{Y}_1'P'P\,\hat{Y}_1/T) & (\hat{Y}_1'P'P\,X_1/T) \\ (X_1'P'P\,\hat{Y}_1/T) & (X_1'P'P\,X_1/T) \end{pmatrix}$$

$$=\operatorname{plim}\begin{pmatrix} (\hat{\Pi}_1'\dot{X}'\dot{X}\hat{\Pi}_1/T) & (\hat{\Pi}_1'\dot{X}'\dot{X}_1/T) \\ (\dot{X}_1'\dot{X}\hat{\Pi}_1/T) & (\dot{X}_1'\dot{X}_1/T) \end{pmatrix} = \begin{pmatrix} \Pi_1'\dot{\Psi}\,\Pi_1 & \Pi_1'\dot{\Psi}_1 \\ \dot{\Psi}_1'\Pi_1 & \dot{\Psi}_{11} \end{pmatrix} = \Delta^{-1} \qquad (34)$$

where $\dot{\Psi}$, $\dot{\Psi}_1$ and $\dot{\Psi}_{11}$ are defined in equations (9), (10) and (11). Hence

$$\sqrt{T}(\hat{\delta}_{1V}-\delta_1) \sim N(0,\ \sigma_{11}\Delta\,\Delta^{-1}\Delta) \sim N(0,\ \sigma_{11}\Delta) \qquad (35)$$

where Δ is consistently estimated by $(W_2'\dot{Z}_1)^{-1}$.

Theil(II) Estimator

The difference between Theil's original estimator and its second version is the order in which the two problems, simultaneity and autocorrelation, are tackled. We have presented the second version in terms of instrumental variables but it is possible to obtain an asymptotically

equivalent estimator by folowing the usual 2SLS approach and replacing the right hand endogenous variables by their estimated values, transforming the equation by P and then applying OLS. We will call this estimator Theil(II).

First write the reduced form (3) as

$$\hat{Y}_1 = X\,\hat{\Pi}_1$$

where

$$\hat{\Pi}_1 = (X'X)^{-1}X'Y_1 \tag{36}$$

Therefore, the transformed fitted values of Y_1 will be

$$\dot{\hat{Y}} = P\hat{Y} = PX(X'X)^{-1}X'Y_1 \tag{37}$$

Thus the Theil(II) estimator is

$$\hat{\delta}_1 = \begin{pmatrix} \dot{\hat{Y}}_1{}'\dot{\hat{Y}}_1 & \dot{\hat{Y}}_1{}'\dot{X}_1 \\ \\ \dot{X}_1{}'\dot{\hat{Y}}_1 & \dot{X}_1{}'\dot{X}_1 \end{pmatrix}^{-1} \begin{pmatrix} \dot{\hat{Y}}_1{}' \\ \\ \dot{X}_1{}' \end{pmatrix} \cdot \quad \dot{y}_1 \tag{38}$$

Consistency of this estimator can be proved as follows

Since $Y_1 = \hat{Y}_1 + \hat{V}_1$

where

$$\hat{Y}_1 = X(X'X)^{-1}X'Y_1$$

the structural equation to be estimated can be written as

$$Py_1 = P\hat{Y}_1\beta_1 + PX_1\gamma_1 + Pu_1 + P\hat{V}_1\beta_1$$

or

$$\dot{y}_1 = \dot{\hat{Y}}_1\beta_1 + \dot{X}_1\gamma_1 + \dot{u}_1 + \dot{\hat{V}}_1\beta_1 \tag{39}$$

Equation (39) can be written as

$$\dot{y}_1 = \dot{W}_2\delta_1 + \dot{w}_1$$

where

$$\dot{W}_2 = (\dot{\hat{Y}}_1 \quad \dot{X}_1)$$

$$\dot{w}_1 = \dot{u}_1 + \dot{\hat{V}}_1\beta_1$$

Therefore, the estimator of δ_1 using OLS will be

$$\hat{\delta}_1 = (\dot{W}_2'\dot{W}_2)^{-1}\dot{W}_2'\dot{y}_1 = (\dot{W}_2'\dot{W}_2)^{-1}\dot{W}_2'(\dot{Z}_1\delta_1 + \dot{u}_1)$$

$$= (\dot{W}_2'\dot{W}_2)^{-1}\dot{W}_2'\dot{Z}_1\delta_1 + (\dot{W}_2'\dot{W}_2)^{-1}\dot{W}_2'\dot{u}_1 \qquad (40)$$

For consistency to hold, we need to show that the probability limit of the first term on the right hand side of equation (40) is equal to δ_1 and that of the second term is equal to zero. From equations (31) and (34) we have

$$\text{Plim}(\dot{W}_2'\dot{W}_2/T) = \Delta^{-1} \qquad (41)$$

$$\text{Plim}(\dot{W}_2'\dot{Z}_1/T) = \Delta^{-1} \qquad (42)$$

To prove the orthogonality of $\dot{W}_2$ and $\dot{u}_1$ note that

$$\text{plim}(\dot{W}_2'\dot{u}_1/T) = \text{plim}\begin{pmatrix} (\hat{Y}_1'P'P\ u_1/T) \\ \\ (X_1'P'P\ u_1/T) \end{pmatrix} = 0 \qquad (43)$$

Since $\hat{Y}_1$ and X_1 are orthogonal to u_1.

Therefore, taking the probability limit of equation (40) using the results in (41), (42), and (43) will prove the consistency of the Theil(II) estimator,i.e.

$$\text{plim}\ (\hat{\delta}_1) = (\Delta)^{-1}(\Delta)\delta_1 + (\Delta)\ \text{plim}(\dot{W}_2'\dot{u}_1) = \delta_1$$

To evaluate the asymptotic efficiency of the Theil(II) estimator, we shall show, following Buse(1983), that its asymptotic distribution is the same as the Theil(IV) estimator and therefore is equal to that of the Theil's G2SLS estimator. To demonstrate this, write the Theil(IV) estimator as

$$(\dot{W}_2'\dot{Z}_1)\ \hat{\delta}_{1V} = \dot{W}_2'\dot{y}_1$$

Substituting for $\dot{Z}_1$ by

$$\dot{Z}_1=(\dot{Y}_1 \ , \ \dot{X}_1) = (\dot{\hat{Y}}_1+\dot{\hat{V}}_1 \ , \ X_1)= \dot{W}_2+(\hat{V}_1 \ , \ 0) = \dot{W}_2 + A$$

the Theil(IV) estimator will become

$$(\dot{W}_2'\dot{W}_2+\dot{W}_2'A)\ \hat{\delta}_{IV} = \dot{W}_2'y_1 \tag{44}$$

We can write the Theil(II) estimator as

$$(\dot{W}_2'\dot{W}_2)\ \hat{\delta}_1 = \dot{W}_2'y_1 \tag{45}$$

Subtracting (45) from (44) and re-arranging the terms, we will have

$$\sqrt{T}\ (\hat{\delta}_{IV} - \delta_1) - \sqrt{T}\ (\hat{\delta}_1 - \delta_1) = -(T^{-1}\dot{W}_2'\dot{W}_2)^{-1}(T^{-1}\dot{W}_2'A)\sqrt{T}\ \hat{\delta}_{IV} \tag{46}$$

We showed that $\text{plim}(T^{-1}\dot{W}_2'\dot{W}_2) = \Delta^{-1}$. The second term on the right hand side of equation (46) is equal to

$$\text{Plim}\ (T^{-1}\dot{W}_2'A) = \text{Plim}\ T^{-1}\begin{pmatrix} \hat{Y}_1'P'P\ \hat{V}_1 & 0 \\ & \\ X_1'P'P\ \hat{V}_1 & 0 \end{pmatrix} = 0$$

Given the assumptions made above we have

$$\text{Plim}\ (\hat{Y}_1'P'P\ \hat{V}_1/T) = \text{Plim}\ (X_1'P'P\ \hat{V}_1) = 0$$

Since $\sqrt{T}\ \hat{\delta}_{IV}$ has a well defined distribution the right hand side of equation (46) converges asymptotically to zero. In other words, the asymptotic distribution of the Theil(II) estimator is equal to Theil(IV) estimator and hence equal to the Theil G2SLS estimator. However, the small sample characteristics of Theil(II) and Theil(IV) estimators will be different from that of the Theil's G2SLS estimator due to the presence of the additional terms in (31) and (42) that only asymptotically converge to zero.

To summarize, we have shown that the Theil type estimators that use a consistent estimate of the ordinary reduced form coefficients are asymptotically equivalent.

That is to say the order in which the transformation to the standard form and transformation for orthogonality are handled does not affect the asymptotic properties.

The Generalized LIML Estimator

To derive the asymptotic properties of GLIML, we must first differentiate equation (31) of Chapter II with respect to $\beta_1{*} = (1\ ,\ -\beta_1{}')'$ and γ_1 for a given value of 'r' and set the result equal to zero to obtain estimates of the coefficients. This process will result in the following equations

$$(\partial/\partial\beta_1{*})\ [(\beta_1{*}'\pi_2\ \beta_1{*})/\beta_1{*}'\pi_1\ \beta_1{*})] = 0 \tag{47}$$

$$\hat{\gamma}_1 = (\dot{X}_1{}'\dot{X}_1)^{-1}\dot{X}_1{}'Y_1{*}\ \beta_1{*} \tag{48}$$

where

$$\pi_1 = \dot{Y}_1{*}'(I-\dot{X}_1(\dot{X}_1{}'\dot{X}_1)^{-1}\dot{X}_1{}')\dot{Y}_1{*}$$

$$\pi_2 = \dot{Y}_1{*}'(I-\dot{X}(\dot{X}'\dot{X})^{-1}\dot{X})\dot{Y}_1{*}$$

where $\dot{Y}_1{*} = (\dot{y}_1\ ,\ \dot{Y}_1)$

Equations (47) and (48) can be combined, noting that $\beta_1{*} = (1\ -\beta_1{}')'$, to give

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{pmatrix} = \begin{pmatrix} \dot{Y}_1{}'\dot{Y}_1 - \lambda\dot{Y}_1{}'(I-\dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}')\dot{Y}_1 & \dot{Y}_1{}'\dot{X}_1 \\ \dot{X}_1{}'\dot{Y}_1 & \dot{X}_1{}'\dot{X}_1 \end{pmatrix}^{-1} \begin{pmatrix} \dot{Y}_1 - \lambda\dot{Y}_1(I-\dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}') \\ \dot{X}_1{}' \end{pmatrix} \dot{y}_1 \tag{49}$$

where λ is equal to

$$\lambda = \mathrm{Min}\ [(\beta_1{*}\ \pi_1\ \beta_1{*}')/(\beta_1{*}\ \pi_2\ \beta_1{*}')]$$

Equivalently, λ is the smallest root of the determinantal equation

$$|\pi_1 - \lambda\ \pi_2| = 0$$

It is well known that LIML estimator is a member of the K-class estimators with $K=\lambda$. Moreover, it can be shown that[29]

$$\text{plim}(\hat{\lambda} - 1) = 0$$

To demonstrate consistency we take the probability limit of the equation (49), noting that $\text{plim}(\lambda)=1$, and substituting

$$\dot{y}_1 = \dot{Y}_1\beta_1 + \dot{X}_1\gamma_1 + \dot{u}_1$$

to obtain

$$\text{plim}\begin{pmatrix}\hat{\beta}_1 \\ \\ \hat{\gamma}_1\end{pmatrix} = \begin{pmatrix}\beta_1 \\ \\ \gamma_1\end{pmatrix} + \Delta\ \text{plim}\ T^{-1}\begin{pmatrix}\dot{Y}_1{}'\dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}' \\ \\ \dot{X}_1{}'\end{pmatrix}\epsilon_t \qquad (50)$$

where Δ stands for the asymptotic covariance matrix of the Theil's G2SLS estimator.

GLIML is consistent if the second term on the right hand side of the above equation is asymptotically equal to zero. The second term can be written as

$$\Delta^{-1}\text{plim}\ T^{-1}\begin{pmatrix}\dot{X}\ \hat{\Pi}_1 \\ \\ \dot{X}_1\end{pmatrix}'\epsilon_t \qquad (51)$$

Since by assumption X's are uncorrelated with the ϵ_t, the above term will be equal to zero and thus GLIML is consistent.

Since $\text{plim}\ \hat{\lambda} = 1$, it follows from (49) that GLIML esimator is identical to G2SLS in the limit. Therefore, GLIML will have the same asymptotic efficiency as the G2SLS.

---

[29]See Goldberger(1964,1965).

IV. METHODOLOGY OF RESEARCH AND CRITERIA OF EVALUATION

Most of the knowledge about econometric estimators concerns their large sample properties. However, an econometric practitioner typically deals with small samples. Therefore, it is of great interest to inquire into the small sample properties of these estimators. The relative lack of knowledge about the small sample performance of estimators stems from the inherent difficulties in their analytical derivation. This fact is the main reason for resorting to numerical analysis or computer simulation as an alternative to mathematical or analytical derivation. Naylor(1971, p.2) defines simulation as

> a numerical technique for conducting experiments with certain types of mathematical models which describe the behavior of a complex system on a digital computer over extended periods of time.

Monte Carlo simulation is a technique of performing sampling experiments on a model using random or pseudorandom numbers. It is a capital intensive (Summers, 1965) approach to the problem of assessing the properties of estimators when analytical or labor-intensive methods become either very complex or impossible.

However, the results of the Monte Carlo simulation cannot be treated without qualification. As Rubinstein(1981, p.10) explains

> Simulation is indeed an invaluable and very versatile tool in those problems where analytical techniques are inadequate. However, it is by no means ideal. Simulation is an imprecise technique. It provides only statistical rather than exact results, and only compares alternatives rather than generating the optimal one....

The inherent imprecision of the Monte Carlo method is partly the result of the sampling error that can be high depending on the number of replications. This defect is mainly due to cost constraints that restrict the number of replications per experiment. Moreover, traditionally, Monte Carlo studies have used the direct simulation method that often produced results that were indeterminate and sometimes contradictory.

These problems can be partially resolved by using more efficient techniques than the conventional direct simulation method discussed by Dhrymes(1970) or Smith(1973). In fact, numerous techniques have been proposed[30] for increasing the efficiency of Monte Carlo simulations over and above what can be obtained by conventional or direct simulation. In the following we discuss the three main Monte Carlo techniques used in econometric research.

A. Direct Simulation or Crude Monte Carlo

Let $\epsilon_1,\ldots,\epsilon_n$ be a sample of N independent random numbers rectangularly distributed between zero and one.[31] Define G_i as a function of ϵ's such that

$$G_i = f(\epsilon_i) \qquad i=1,\ldots,N \tag{1}$$

Therefore, the quantities G_i are also independent random

----

[30]See Naylor(1966,1971), Hammersley and Handscomb(1964), Rubinstein(1981).

[31]Note that frequently the random normal numbers are derived from rectangularly distributed random numbers lying in the interval (0,1) using appropriate transformations. As we shall discuss later, the random normal number generator used in this study also uses this method.

variables with mean equal to $\overline{G}$ and variance of $\sigma^2(\overline{G})$. Then, an unbiased estimator of G can be obtained as

$$\hat{G} = (\sum_{i}^{N} G_i)/ N \tag{2}$$

The variance of $\hat{G}$ will be equal to

$$\mathrm{Var}(\hat{G}) = \sigma^2(\overline{G})/ N \tag{3}$$

In practice, $\sigma^2(\overline{G})$ is usually unknown and has to be estimated. Using the sample variance of G_i, the unbiased estimator of $\sigma^2(\overline{G})$ will be equal to

$$S^2(G_i) = \sum_{i}^{N}(G_i-\hat{G})^2/(N-1) \tag{4}$$

Therefore, the unbiased estimator of the variance of $\hat{G}$ is

$$\mathrm{Var}(\hat{G}) = S^2(G_i)/N = [1/N(N-1)] \sum_{i}^{N}(G_i-\hat{G})^2 \tag{5}$$

$\hat{G}$ is referred to as a "Crude Monte Carlo" estimate of $\overline{G}$.

In direct simulation, as equation (3) indicates, cutting down the standard error by a factor of μ requires increasing the number of replications by a factor of μ^2. This number can be drastically cut down by using more efficient simulation techniques.

In direct Monte Carlo simulation, no refinement in the choice of the random numbers are being made. Variance reducing techniques can be viewed as the methods which use known information about the problem (Rubinstein, p.121). Two of these variance reduction techniques are control variates and antithetic variates.

B. Control Variate Technique

The sampling error in the direct or crude Monte Carlo simulation of $\overline{G}$ arises from the variation of $(G_i - \overline{G})$ over replications as ϵ_i runs over $0<\epsilon_i<1$. The control variates technique is a method which reduces this variation and therefore improves the efficiency of simulation.

Mikhail(1972,1975) applied the control variate technique for two stage and three stage least squares and full information maximum likelihood for static models and achieved considerable efficiency gains in estimating means and variances. Hendry and Harrison(1974) employed control variates to estimate the biases of the OLS and 2SLS for various specification of a dynamic structure with autocorrelated errors and also obtained significant efficiency gains.[32]

The essence of the control variate technique is to find an auxiliary statistic C such that $\hat{G}$ and C are positively correlated and the distribution of C is known. Using C we control as much as possible of the variation of $\hat{G}$ analytically and estimate the remainder. Through this approach we minimize the variation of the parameters to be estimated by absorbing most of their variations using control variates. Hence, the resulting Monte Carlo errors will be considerably reduced.

Consider C to be a control variate for $\hat{G}$ with known expectation E(C) and variance Var(C). Then, we use C to

----

[32] See also Hendry(1979), Hendry and Srba(1977).

construct an estimator for $\overline{G}$ with a smaller variance than the estimator $\hat{G}$. Therefore, instead of investigating the direct simulation estimator $\hat{G}$, one can use

$$\hat{G}* = \hat{G} - \Theta(C- E(C)) \tag{6}$$

where Θ is a parameter that will be determined shortly. From (6) we have

$$E(\hat{G}*)=E(\hat{G})$$

which is an unbiased estimator of $\overline{G}$. Moreover, the variance of $\hat{G}*$ will be

$$Var(\hat{G}*) = Var(\hat{G}) + \Theta^2 Var(C) - 2\Theta Cov(\hat{G},C) \tag{7}$$

which will be smaller than $Var(\hat{G})$ if

$$\Theta^2 Var(C) < 2\Theta Cov(\hat{G},C) \tag{8}$$

Therefore, the value of Θ can be found in such a way to minimize the $Var(\hat{G}*)$. This value will be equal to

$$\hat{\Theta} = [Cov(\hat{G},C)/ Var(C)] \tag{9}$$

Substituting $\hat{\Theta}$ in (7) yields

$$Var(\hat{G}*) = (1-R_k{}^2)\ Var(\hat{G}) \tag{10}$$

Where R_k is the correlation coefficient between $\hat{G}$ and C. Thus the higher the R_k, the smaller the $Var(\hat{G}*)$ and therefore the greater the reduction in variance.

Equation (10) provides an important intuitive explanation of the control variate technique. It demonstrates that if nothing is known about the estimator $\hat{G}$, then no part of that estimator can be solved analytically, i.e., $R_k=0$, and therefore the variance of $\hat{G}*$ will be equal to $Var(\hat{G})$. In other words, there will be no reduction in variance and the control variate technique is not

applicable. As information about $\hat{G}$ increases a greater part of its variation can be solved analytically through C and thus R_k increases which in turn decreases the variance of $\hat{G}*$. If everything is known about $\hat{G}$, its whole variations can be explained analytically and R_k becomes equal to one. This means that in the case of perfect knowledge, the variance of the estimator $\hat{G}*$ is equal to zero.

The main problem for application of this method is to find a C that is highly correlated with $\hat{G}$ and whose distribution can be analytically derived. In practice, it is usually the lack of knowledge about the distribution of $\hat{G}$ which necessitates the experiment. Therefore, it is highly unlikely to find an estimator C whose distribution is close to $\hat{G}$.

C. Antithetic Variates Technique

The antithetic variates technique is another method for reducing the sampling variation of a Monte Carlo experiment. It is due to Hammersley and Morton(1956).[33] The basic idea of the Antithetic variates technique is to find another statistic $\hat{G}**$ having the same expectation as $\hat{G}$ and which has a strong negative correlation with $\hat{G}$. Then the combination of these two will result in a estimate with the same expectation as $\hat{G}$ but smaller variance. It can be viewed as a stratification of the range of the random errors into two opposite (negatively related) sections and then sampling

---

[33] See also Hammersley and Handscomb(1964), Rubinstein(1981), Hammersley and Mauldon(1956).

equally from each half.[34] Therefore, the two sets of estimators corresponding to two negatively related sections would mutually compensate each other's variations without affecting the unbiasedness of the result.

This method was employed by Mikhail(1972,1975). Hendry and Trivedi(1972) also used this method in a study of maximum likelihood estimation of difference equations with moving average errors. The results were a considerable gain in efficiency in estimating the variance and reducing the small sample bias. Hendry and Trivedi reported gains in efficiency of about 300 to 500 percent in estimation of the biases in the structural parameters.

The antithetic method proceeds as follows:

Let $\hat{G}{**}$ be the antithetic variate for $\hat{G}$, then define

$\hat{G}* = (\hat{G}{**} + \hat{G})/2$

which will be an unbiased estimator of $\bar{G}$, and its sampling variance will be equal to

$$\mathrm{Var}(\hat{G}*) = 1/4\ \mathrm{Var}(\hat{G})+1/4\ \mathrm{Var}(\hat{G}{**})+1/2\ \mathrm{Cov}(\hat{G}{**},\ \hat{G}) \qquad (11)$$

If the cov($\hat{G}{**}$, $\hat{G}$) can be made strongly negative, then the sampling variance of $\hat{G}*$ will be drastically reduced below the random sampling outcome, Var($\hat{G}$). For the choice of antithetic variate, recall that $G_i = f(\epsilon_i)$, where the ϵ_i's were independently rectangularly distributed random numbers between zero and one. One can use $\epsilon_i* = 1-\epsilon_i$ to get an antithetic estimator $\hat{G}{**}$ which is likely to be highly negatively correlated with $\hat{G}$. The random numbers ϵ_i* will

----

[34] See Hendry and Harrison,1974, p.156.

also be rectangularly distributed in the range zero and one. Furthermore, random normal variables corresponding to ϵ_i* are frequently derived using the transformation[35]

$$U_i* = \sum_{1}^{12} \epsilon_i* - 6 \qquad (12)$$

These random normal variables will have the same mean and variance as the random normal variables corresponding to ϵ_i. Moreover, using the above transformation, the random normal variables corresponding to ϵ_i will be equal in magnitude to the ones corresponding to ϵ_i* with opposite sign.

The relative efficiency of two Monte Carlo methods can be defined as follows. Assume two Monte Carlo methods require T_1 and T_2 units of computer time or number of replications, respectively. Also assume that the resulting estimates of the parameter under consideration have sampling variances σ_1^2 and σ_2^2. Method one is said to be more efficient if

$$E = (T_1\sigma_1^2)/(T_2\sigma_2^2) < 1$$

Clearly, using the antithetic variate method requires doubling of the number of replications relative to the one required by the crude Monte Carlo method. Therefore, the antithetic estimator $\hat{G}*$ will be more efficient than the direct simulation only if

----

[35]We used the REGEN program developed by Haitovsky and Jacobs(1972) to generate random normal variables. To generate the uniform random numbers REGEN uses the RANNO (Harvard Computing Center) program which utilizes the Power Residue Method to generate random numbers between zero and one. To produce the normal random numbers REGEN employs subroutine GAUS, which applies the Central Limit theorem to 12 uniform random numbers obtained from RANNO to generate one normal number.

$$\mathrm{Var}(\hat{G}*) < 1/2\ \mathrm{Var}(\hat{G}) \qquad (13)$$

Rubinstein(1981, pp.135-6) proved that under the condition that $G(\epsilon_i)$ is a continuous monotonically non-increasing (non-decreasing) function with continuous first derivatives, using ϵ_i and ϵ_i* as antithetic variates will always satisfy the inequality (13).

The merit of the antithetic variate technique lies in its simplicity. AS Hammersley and Handscomb argued(1964, p.61)

> From the practical viewpoint, the mathematical conditions that a Monte Carlo technique has to satisfy govern its efficiency. As in the case of importance sampling and control variates, we are usually unable to satisfy the conditions in the theoretically optimum way, and we have to be content with some compromise. When the conditions are fairly loose and flexible, it is easier to reach a good compromise. This is the case with antithetic variates; in practice it is relatively easy to find a negatively correlated unbiased estimator of G, usually easier than it is to find an equally satisfactory control variate or importance function. Accordingly, the antithetic variate method tends to be more efficient in practice.

In the present work we shall employ both the direct Monte Carlo technique and the antithetic variate method in simulating the small sample properties of econometric estimators.

D. Criteria of Evaluation

To compare the finite sample properties of estimators, there is a need for some measures by which the performance of each estimator can be evaluated. These measures or the criterion of goodness represent the summary characteristics

of the sampling distribution of the estimator.

If $\hat{G}_{ij}$ represents the direct simulation or the result of the antithetic variate method of the estimate of the jth parameter G_j in the ith replication, then the evaluation criteria used in this study are the following:

1. The mean of the sample estimate

$$\hat{\bar{G}}_j = \sum_{i=1}^{N} \hat{G}_{ij} / N \tag{14}$$

2. The bias of $\hat{G}_{ij}$ is equal to $\hat{G}_{ij}-G_j$, therefore the mean bias of the sample estimate is

$$\text{BIAS}(j) = \sum_{i=1}^{N} (\hat{G}_{ij}-G_j) / N = \hat{\bar{G}}_j - G_j \tag{15}$$

3. The Root Mean Square Error (RMSE) which takes into account bias and dispersion of the estimates at the same time and represents the discrepancy between the sample means and the corresponding true parameter values; it is defined as

$$\text{RMSE}(j) = [\sum_{i=1}^{N} (\hat{G}_{ij} - G_j)^2 / N]^{1/2} \tag{16}$$

4. Defining the Mean Squared Error(MSE) matrix as

$$\text{MSE} = [\sum_{i=1}^{N} (\hat{G}_i - G)'(\hat{G}_i - G)]/N \tag{17}$$

where N is the number of replications and G_i is a K.1 vector whose jth coefficient is G_{ij}. Trace(MSE) can be defined as

$$\text{Trace(MSE)} = \sum_{j=1}^{K} \text{diag(MSE)} \tag{18}$$

where K is the number of parameters to be estimated.

The Trace(MSE) index provides a aggregate measure used to compare the relative efficiency of different estimators.

5. A good summary statistic of the aggregate bias of the

estimated coefficients is the Euclidean distance. If there are K parameters in the equation under consideration, the Euclidean distance is equal to

$$d = [\sum_{j=1}^{K}(\hat{\bar{G}}_j - G_j)^2]^{1/2} \qquad (19)$$

6. Finally, a good summary index of the aggregate dispersion is the determinant of the Mean Squared Errors(MSE) matrix suggested by Dhrymes(1971). Using det(MSE), the relative efficiency of two methods can be measured by

$$e = \det(MSE_1)/ \det(MSE_2) \qquad (19)$$

If $e<1$, we can conclude that method one is more efficient than method two.

In using the above statistics we implicitly assumed the existence of the first two moments of the estimators' distribution. Since the estimators are the ratio of sums of random variables, it is a possibility that their moments are not well behaved. For instance, the existence of a number of extremely aberrant parameter estimates can cause the variance to tend to infinity. In these circumstances the comparison of estimators based on the moments which do not exist is wrong and potentially misleading.

This problem was first discussed by Basmann (1960,1961) and Nagar(1959,1960). Nagar worked in a very general context. He is mainly concerned with the approximate distributions of K-class and double K-class estimators. Basmann's work is more specific and primarily deals with the exact sampling distributions of several two stage least

squares estimators. He demonstrates many instances where the distributions of the estimators do not possess a finite mean or variance. However, as we shall discuss later, the first two moments of the generalized classical linear estimators exist for the models considered in this study.

Apart from the comparison of the efficiency of the estimators on the basis of the above mentioned statistics, an econometric practitioner is likely to be interested in the reliability of the estimators for hypothesis testing. Park and Mitchell(1980) studied the performance of the different estimators in hypothesis testing when the classical linear model had autocorrelated disturbances and the explanatory variables were trended. They found that all the estimators they considered seriously underestimate standard errors and made estimated coefficients appear to be much more significant than they actually were. This study will attempt to assess whether their results carry over to simultaneous equation systems.

V. EXPERIMENTAL DESIGN AND RESEARCH STRATEGY

In order to investigate the small sample properties of the limited information estimators, the following three-equation model, which is the same as the one used by Cragg(1967a,1967b,1968), was employed[36]

$$Y\,B + X\,\Gamma = U$$

$$U = U_{-1}R + E$$

where Y is a T.3 observation matrix on the endogenous variables, X is a T.7 observation matrix of the exogenous variables whose first element at every observation is unity, U is the T.3 matrix of the disturbances, R is a 3.3 diagonal matrix of the autocorrelation coefficients and E is the T.3 matrix of random errors. B and Γ are 3.3 and 7.3 matrices of structural coefficients to be estimated.

The B and Γ matrices that are used in all the experiments in this part are[37]

$$B' = \begin{pmatrix} 1.00 & -0.64 & -0.22 \\ -0.72 & 1.00 & 0.00 \\ 0.00 & -0.17 & 1.00 \end{pmatrix}$$

----

[36] Most of the Monte Carlo studies have used a two equation model. In his study Mosbaek(1970) argued that "the three equation size was selected because pilot Monte Carlo runs showed that they revealed the essential properties of larger models whereas two equation models do not".

[37] Cragg(1967a,1967b) used five different sets of structural parameters. His second set is employed in this study. The reason for the choice of structure two was that it was the only structure that did not result in occurrence of any singular or apparently singular matrices or "nonsensical" results. See Cragg(1967a), Table 3, p.82.

$$\Gamma' = -\begin{pmatrix} 32 & 0.65 & 0.00 & 0.00 & 0.52 & 0.00 & 0.00 \\ 18 & 0.00 & 0.68 & 0.00 & 0.67 & 0.00 & 0.34 \\ 37 & 0.00 & 0.95 & 0.39 & 0.00 & 0.18 & 0.00 \end{pmatrix}$$

The number of a priori restrictions in terms of excluded exogenous variables in each equation is greater than the number of endogenous variables included in the same equation. This implies that the Basmann(1960,1961) conditions for the existence of the first two finite sample moments are satisfied.[38]

As was mentioned in Chapter I, the statistical characteristics of different estimators are likely to be sensitive to the specification of the process that generates the exogenous variables. Specifically, the recent Monte Carlo studies by Maeshiro (1976,1979), Park and Mitchell(1981) and Beach and Mackinnon(1978) have emphasized the effect of trended data on the estimators which omit the first observation. In general, Monte Carlo studies have used three types of specifications for the exogenous variables; namely

1. A stationary autoregressive process of the form[39]

----

[38]Note that the Basmann results were derived for models without autocorrelation. The presence of autocorrelation , when the model does not contain any lagged endogenous variables, does not create any additional problem for identification. In fact, using additional information about the structure of the disturbances would aid identification. Therefore, since the Basmann results depend on the degree of overidentification of the structural equation, we can expect them to hold true for the present case. For discussion of identification and autocorrelation, see F.M. Fisher(1966), M.Deistler (1975,76), M.Deistler and J.Schrader (1979), C.Hsiao (1981).

[39]This specification was used by Rao and Griliches(1969) and Spitzer(1979).

$$X_t = \lambda X_{t-1} + v_t$$

where

$$v_t \sim N(0, \sigma_v^2)$$

2. A non-stochastic autoregressive process such as[40]

$$X_t = \lambda X_{t-1} = \lambda^t X_0$$

3. A stochastic trended process as[41]

$$X_t = \exp(\lambda t) + w_t$$

where

$$w_t \sim N(0, \sigma_w^2)$$

We have employed all three types of specification.[42] We have specifically examined Maeshiro's and Park and Mitchell's conjecture that the results of the Monte Carlo studies which have not used trended data (i.e., that of Rao and Griliches) are questionable. For this purpose we have examined a number of variations of trended stochastic and non-stochastic specifications (case 2 and case 3). In particular, we have examined cases in which trend exist in most of the exogenous variables of the simultaneous equation system and cases in which the trended variables appear only in the structural equation under consideration. The primary reason for this choice is that some estimators such as Fair employ only the exogenous variables that are included in the structural equation under consideration as instruments. Other estimators such as Theil use all the explanatory variables

----

[40] This specification was used by Maeshiro(1976,1979).
[41] This specification was used by Beach and Mackinnon(1978).
[42] All the computations were carried out on the university of Alberta computer facilities. The random numbers were generated using Regen-Computer Program developed by Haitovsky and Jacobs(1972).

as instruments. Therefore, the existence of trend in the exogenous variables that are excluded from the structural equation under consideration is likely to affect the latter type of estimators more seriously.

To minimize the probability of exclusion of any potentially good estimator, a pilot experiment which covers all the possible variations of different estimators was undertaken. The results of the pilot run was carefully analysed and the estimators or the variations of them which performed poorly relative to other estimators were excluded from the subsequent experiments. This strategy had the following advantages: 1- It covered all potential estimators; 2- It minimized the total cost of the study; 3-It minimized the amount of statistics to be analysed and thus makes the results more comprehensible.

The pilot experiment undertaken in this part investigates the performance of 13 estimators when the sample size is equal to 30, the autocorrelation coefficient is equal to 0.6 and the exogenous variables are specified to be trended and non-trended.

The disturbance terms, E_t, are jointly normally distributed with a zero mean and variance-covariance matrix of the form

$$\Theta = E(E_t E_t') = \begin{pmatrix} 22.00 & 7.13 & 15.24 \\ 7.13 & 20.00 & 16.77 \\ 15.24 & 16.77 & 25.00 \end{pmatrix}$$

Each complete experiment consists of 200 replications, 100 "thetic" and 100 antithetic. Given the covariance structure Θ, a set of E_{it} ($i=1,2,3$, $t=1,\ldots,T$) was generated. Given the structure of the X matrix and the values of the structural parameters (B's, Γ's, r's) the endogenous variables were calculated and the structural parameters were estimated. These estimated coefficients were stored as thetic estimates. The E_{it}'s were then replaced by $(-E_{it})$ and new U_t and Y_t were generated. The structural parameters were estimated again by the same estimators and stored as "antithetic" estimates. The statistics based on the direct simulation and the antithetic method are then reported. The summary table of the 13 estimators employed in the pilot study for the estimation of the structural coefficients of the first equation of the above model and their computer programs are presented in appendix II.[43]

Table 5-1 categorizes the estimators on the basis of the reduced form they use and the number of observations they utilize. Estimators on the same row of Table 5.1 are basically the same except for the reduced form they employ in their first stage estimation and the number of observations (T or T-1) they utilize and thus are directly comparable.

----

[43] All the computer programing was done in the APL language using University of Alberta computer facilities.

TABLE 5-1: Structure of the Estimators

Augmented Reduced Form	Ordinary Reduced Form	
T-1	T	T-1
Fair		
	Theil(CORC)	MG2SLS
	Theil(P.W)	
	Theil(IV)	
	Theil(CUBIC)	
	Theil(II)	
ALIML	GLIML	GLIML
FBrundy	MBrundy	MBrundy

Note that T is the number of observations used by each method.

A. Monte Carlo Results of the Pilot Experiment

The following chart shows the structure of the pilot study.

Autocorrelation of the Structural Disturbances	No. of Obs.	Data Specification
$r_1=0.6$	N=30	Stochastic Non-trended
$r_2=0.9$	N=30	Stochastic Trended
$r_3=0.2$	N=30	Non-Stochastic Trended

Relative Efficiency of Different Monte Carlo Methods

Turning now to the results, Table 5-2 presents the mean bias and its standard error for the Fair and Theil's G2SLS estimator.[44] The relative efficiency of the two methods (defined as the ratio of the variances of the estimated coefficients weighted by the number of replications for both estimators) is shown in the last two columns of Table 5-2. The antithetic estimates are based on twice as many replications as the direct simulation estimates. However, it can be seen that generally the antithetic method has resulted in gains in efficiency over and above what could be achieved by employing the direct simulation method with twice as many replications. In general, we achieved an average gain in efficiency of about 15 to 28 percent for the

---

[44] The exogenous variables were stochastic non-trended as defined in the next section.

TABLE 5-2: Relative Efficiency of Different Methods

	Fair				Theil(G2SLS)				$T_1\sigma_1^2 / T_2\sigma_2^2$	
	Direct Simulation		Antithetic Variates		Direct Simulation		Antithetic Variates		Fair	Theil
Coef	Bias	S.E	Bias	S.E	Bias	S.E	Bias	S.E		
C	-15.55	2.47	-16.74	1.94	-1.50	1.37	-3.13	0.62	1.23	0.41
B_1	0.24	0.025	0.21	0.02	0.12	0.02	0.11	0.01	0.92	0.50
B_2	-0.16	0.035	-0.10	0.02	-0.17	0.03	-0.13	0.02	0.65	0.89
C_1	-0.16	0.02	-0.14	0.01	-0.13	0.02	-0.09	0.01	0.50	0.50
C_2	-0.54	0.06	-0.51	0.04	-0.24	0.04	-0.21	0.03	0.89	1.12
r	-0.22	0.03	-0.21	0.02	-0.13	0.03	-0.12	0.02	0.89	0.89

Note that in all the Tables in this chapter the structural equation under consideration is

$$y_1 = C + Y_1 B_1 + Y_2 B_2 + X_1 C_1 + X_4 C_2 + u_1$$

Fair and Theil estimators, respectively. Therefore, in analysing the results, we focus on the results of the antithetic estimates (Tables 5-3, 5-4).

Analysis of the Results

<u>Stochastic Non-trended Exogenous Variables:</u>

In this part the exogenous variables follow a stationary autoregressive process of the form used by Rao and Griliches with the following specification[45]

	X_1	X_2	X_3	X_4	X_5	X_6
Mean	10	12	14	8	9	11
Standard Error	5.78	6.93	8.10	4.62	5.20	6.36
λ	0.6	0.5	0.4	0.3	0.2	0.1

Evaluating estimators on the basis of their bias and RMSE, one can summarize the results given in Table 5.3 in the following points:

1. As was expected, all estimators under-estimated the autocorrelation coefficient. Those estimators which use the Prais-Winsten method for estimating the autocorrelation coefficient had the lowest bias.
2. Using the Beach and Mackinnon cubic formula instead of the Prais-Winsten method for estimating the

---

[45] The ratio of the mean to the standard error of the exogenous variables are the same as the ones employed by Goldfeld and Quandt(1972) in their study of the small sample properties of the estimators designed for estimation of the simultaneous equation models characterized by autocorrelation. This specification of the exogenous variables will be the same in all the experiments that use stochastic non-trended exogenous variables.

TABLE 5-3: Estimated Bias Using Stochastic Non-Trended Exogenous Variables
(r=0.6, T=30)

	T-1						T						
Coef	Fair	FBrundy	MBrundy	MG2SLS	ALIML	GLIML	Theil (P.W)	Theil (CORC)	Theil (CUBIC)	Theil (IV)	Theil (II)	MBrundy	GLIML
C	-16.74	0.63	-1.30	-3.32	-5.74	-4.67	-3.13	-3.14	-3.15	-2.98	-3.16	-0.11	-4.63
B_1	0.21	-0.07	0.00	0.10	-0.12	0.08	0.11	0.11	0.11	0.13	0.14	-0.14	0.08
B_2	-0.10	0.10	0.00	-0.11	0.11	-0.06	-0.13	-0.13	-0.13	-0.16	-0.17	0.22	-0.07
C_1	-0.14	0.05	0.07	-0.09	0.04	-0.07	-0.09	-0.08	-0.09	-0.10	-0.11	0.16	-0.07
C_2	-0.51	0.09	0.06	-0.21	0.25	-0.19	-0.21	-0.23	-0.21	-0.24	-0.25	0.24	-0.17
r	-0.21	-0.13	-0.14	-0.15	-0.15	-0.27	-0.12	-0.14	-0.14	-0.12	-0.12	-0.11	-0.26
Aggregate* Bias	0.61	0.21	0.17	0.31	0.34	0.35	0.31	0.33	0.32	0.35	0.37	0.40	0.34

*This index does not include the intercept term.

TABLE 5-4: Estimated Root Mean Square Error Using Stochastic Non-Trended Exogenous Variables
(r=0.6, T=30)

	T-1						T						
Coef	Fair	FBrundy	MBrundy	MG2SLS	ALIML	GLIML	Theil (P.W)	Theil (CORC)	Theil (CUBIC)	Theil (IV)	Theil (II)	MBrundy	GLIML
C	25.53	26.10	24.35	8.11	29.36	11.40	6.97	6.99	6.98	10.81	14.23	21.55	10.34
B_1	0.27	0.40	0.90	0.17	0.63	0.30	0.17	0.17	0.17	0.22	0.28	0.69	0.28
B_2	0.24	0.45	1.31	0.22	0.72	0.43	0.23	0.23	0.23	0.29	0.35	1.03	0.40
C_1	0.19	0.23	0.73	0.16	0.51	0.34	0.14	0.16	0.15	0.17	0.23	0.54	0.33
C_2	0.66	0.85	1.44	0.39	1.51	0.61	0.34	0.34	0.34	0.45	0.53	1.12	0.58
r	0.32	0.25	0.28	0.28	0.27	0.38	0.25	0.25	0.25	0.25	0.25	0.26	0.36
Trace(MSE)*	0.70	1.20	5.21	0.33	3.53	0.91	0.28	0.29	0.28	0.43	0.60	3.15	0.81

*This index does not include the intercept term.

autocorrelation coefficient in the Theil estimator did not have any significant effect on the estimated coefficients. This result confirms the findings in the single equation context that omission of the Jacobian of the transformation from the likelihood function would not materially affect the estimated coefficients. Since both methods on average resulted in approximately the same number of iterations (i.e., 4.3), it would seem reasonable to choose the Prais-Winsten formula over the cubic one due to its relative simplicity and lower cost.

3. Comparison of different estimators on the basis of their index of overall bias measured by the Euclidean distance[46] indicates that Brundy and Jorgenson estimator had the lowest and Fair estimator had the highest index of overall bias. In general, the index of overall bias showed that none of the estimators performed very poorly relative to others in estimation of the structural coefficients except the intercept. The Fair estimator seriously under-estimated the constant term.

4. The relative efficiency of employing different reduced forms was investigated by comparing different versions of the Brundy and Jorgenson estimator. Fair-Brundy, which uses the augmented reduced form, had lower RMSE for all the estimated coefficients except the intercept.

----

[46]Note that we have omitted the constant term from calculation of this index. This was mainly due to the fact that the relative large biases in the constant term could seriously distort the overall picture of the performance of the estimators.

However, it must be noted that in two instances the Fair-Brundy estimator resulted in extreme outliers that were not acceptable and were therefore removed from the list of estimated coefficients. Thus, Fair-Brundy estimated results are on the basis of 196 "thetic" and antithetic estimates. Taking into account these two outliers renders the Fair-Brundy estimator inferior to the Brundy estimator.

5. Investigation of the relative efficiency of utilizing the first observation was carried out by comparing Theil G2SLS with the modified generalized two stage least squared estimator, Brundy(T-1) with Brundy(T) and GLIML(T-1) with GLIML(T). Theil and GLIML(T) estimators showed lower RMSE than MG2SLS and GLIML(T-1) for all but one coefficient. Brundy(T) estimator almost invariably dominated Brundy(T-1). These results suggest that using the first observation does increase the small sample efficiency of estimators.
6. The small sample effect of the order in which the two problems of simultaneity and autocorrelation are handled in the Theil type estimators was examined by comparing Theil(G2SLS), Theil(II) and its instrumental variable version Theil(IV). Theil(G2SLS) had the lowest aggregate bias and almost invariably dominated the other two estimators on the RMSE criteria. This result confirms our findings in Chapter III that the second version of the Theil estimator is numerically different from Theil

G2SLS due to the existence of the additional terms that converge asymptotically to zero.

7. Comparison of the ALIML and GLIML suggests that GLIML which is derived under a different set of assumptions performs better in small samples. GLIML showed much lower RMSE for all the estimated coefficients.

Finally, an overall comparison of different estimators, on the basis of the sum of the MSE of the estimated coefficients (excluding the intercept), suggests the superiority of Theil type estimators, followed by MG2SLS, Fair's estimator, the LIML estimators and Brundy Jorgenson type estimators.

Trended Exogenous Variables

In this part we employ two different specifications of trended data, i.e., trended stochastic and trended non-stochastic exogenous variables.

1- Trended stochastic exogenous variables:

To test the effect of trended data on the performance of the limited information estimators we considered two cases: one in which four of the exogenous variable , only one of which (i.e. X_1) appears in the structural equation under consideration, are trended. Then the effect of trend existing only in the structural equation under consideration (i.e., X_1 and X_4) is examined. The exogenous variables in both cases follow the same specification used by Beach and Mackinnon.

Case 1: X_1 and X_2 are the same as in the previous experiment

$$X_3 = e^{0.10t} + w_1 \qquad w_1 \sim N(0,\ 0.0009)$$

$$X_4 = e^{0.11t} + w_2 \qquad w_2 \sim N(0,\ 0.0025)$$

$$X_5 = e^{0.12t} + w_3 \qquad w_3 \sim N(0,\ 0.0049)$$

$$X_6 = e^{0.13t} + w_4 \qquad w_4 \sim N(0,\ 0.0081)$$

The experiments in the single equation context suggest substantial gains in the efficiency of the estimators which employ all T observations relative to the ones which omit the first observation. When the data is trended, however, Tables 5-5 and 5-6 show that this gain in efficiency is not materializing in simultaneous equation systems. These Tables show that the relative efficiency of the first observation for Theil and GLIML(T) and Brundy(T) estimators has increased slightly relative to the non-trended case. Nevertheless, the overall picture has not changed dramatically. The important observation is that Fair's estimator becomes superior to Theil's estimator. This picture was more or less the same when we change the degree of autocorrelation to 0.9 (Table 5-7).

Case 2: In this case only the X's that appear in the structural equation under consideration are trended; the rest are stochastic non-trended as in the case of the first experiment.

$$X_1 = e^{0.15t} + w_1 \qquad w_1 \sim N(0,\ 0.0009)$$

$$X_4 = e^{0.19t} + w_2 \qquad w_2 \sim N(0,\ 0.0025)$$

TABLE 5-5: Estimators Using T-1 Observations and Stochastic Trended Data
(r=0.6, T=30)

	Fair		MG2SLS		Brundy		FBrundy		ALIML		GLIML	
Coefficient	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
C	-8.10	17.88	-3.78	13.23	-4.38	54.03	37.11	329.76	28.61	283.5	-5.14	14.73
B_1	0.26	0.38	0.36	0.44	0.29	2.00	-0.54	4.83	-0.57	7.62	0.37	0.44
B_2	-0.29	0.57	-0.52	0.67	-0.41	3.75	0.26	2.79	0.62	9.52	-0.52	0.67
C_1	-0.19	0.30	-0.29	0.37	-0.29	1.63	-0.18	2.88	0.20	3.11	-0.28	0.36
C_2	-0.66	0.92	-0.72	0.95	-0.55	3.22	1.93	16.99	0.75	10.02	-0.75	0.99
r	-0.26	0.40	-0.46	0.60	-0.46	0.50	-0.27	0.41	-0.34	0.53	-0.35	0.57
Trace(MSE)*		1.57		2.04		31.34		328.24		259.05		2.08

*This index does not include the intercept term.

TABLE 5-6: Estimators Using T Observations and Stochastic Trended Data, r=0.6, T=30 (r=0.6, T=30)

	Theil(P.W)		Theil(IV)		Theil(II)		GLIML(T)		Brundy(T)	
Coefficient	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
C	-3.96	14.10	-2.54	21.42	-3.43	22.91	-3.82	19.34	0.34	55.12
B_1	0.32	0.41	0.29	0.49	0.35	0.52	0.32	0.43	0.22	1.95
B_2	-0.45	0.62	-0.43	0.68	-0.46	0.71	-0.47	0.66	-0.30	3.20
C_1	-0.24	0.33	-0.23	0.34	-0.25	0.37	-0.23	0.36	-0.26	1.60
C_2	-0.62	0.85	-0.55	1.05	-0.59	1.23	-0.64	0.90	-0.50	3.10
r	-0.38	0.54	-0.38	0.53	-0.38	0.55	-0.42	0.67	-0.39	0.53
Trace(MSE)*		1.68		2.20		2.73		2.01		26.49

*This index does not include the intercept term.

TABLE 5-7: Relative Efficiency of Estimators Using Stochastic Trended Data
(r=0.9, T=30)

Coefficient	Fair		MG2SLS		Theil(P.W)	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
C	-8.94	19.72	-10.10	17.96	-11.34	18.88
B_1	0.55	0.67	0.63	0.72	0.64	0.72
B_2	-0.76	1.01	-0.87	1.02	-0.87	1.00
C_1	-0.40	0.50	-0.47	0.56	-0.47	0.55
C_2	-1.06	1.29	-1.28	1.45	-1.27	1.45
r	-0.72	0.81	-0.75	0.83	-0.74	0.83
Trace(MSE)*		4.04		4.66		4.61

*This index does not include the intercept term.

The results summarized in Table 5-8 shows that the Theil estimator dominates Fair and MG2SLS estimators on the basis of trace(MSE) criteria.

2- Trended non-stochastic exogenous variables:

The third set of experiments was conducted with non-stochastic trended data. We considered three cases. One in which the exogenous variables appearing in the structural equation under consideration were downward trended. Then the same experiment was repeated assuming that the exogenous variables were upward trended. Finally, the case in which four of the exogenous variables were trended was considered. The results show that only in the cases where both of the exogenous variables in the equation under consideration were strongly trended, did Theil's estimator perform better than Fair and MG2SLS and this difference was specially noticeable for the coefficients of those two explanatory variables.

Case 1: All X's were stochastic non-trended as in experiment one except X_1 and X_4 which were non-stochastic trended as

$$X_1 = 0.6\ X_{1,-1}$$

$$X_4 = 0.8\ X_{4,-1}$$

Table 5-9 shows the comparison between Fair, Theil and MG2SLS estimators when the autocorrelation coefficient is equal to 0.6. The efficiency of Theil estimator's in estimating the coefficients of X_1 and X_4 is much higher than both the Fair and MG2SLS estimators. Generally, Theil's estimator performed much better than the Fair estimator.

TABLE 5-8: Root Mean Square Error of Different Estimators Using Stochastic Trended Data
(r=0.6, T=30)

Coefficient	Theil	Fair	MG2SLS
C	12.30	25.10	12.93
B_1	0.25	0.41	0.25
B_2	0.36	0.42	0.37
C_1	0.72	0.75	0.73
C_2	0.46	0.73	0.47
r	0.37	0.44	0.38
Trace(MSE)*	1.06	1.63	1.10

*This index does not include the intercept term.

TABLE 5-9: Root Mean Square Error of Different Estimators Using Non-Stochastic Downward Trended Data
(r=0.6, T=30)

Coefficient	Theil	Fair	MG2SLS
C	16.48	34.36	16.41
B_1	0.26	0.61	0.24
B_2	0.35	0.98	0.33
C_1	0.21	0.80	0.82
C_2	0.41	0.97	1.02
r	0.40	0.80	0.47
Trace(MSE)*	0.56	3.55	2.10

*This index does not include the intercept term.

Case 2: In this case the X's that appear in the first equation were upward trended, while the rest of the exogenous variables were left the same as in the first experiment.

$X_1 = 1.10\ X_{1,-1}$

$X_4 = 1.13\ X_{4,-1}$

As Table 5-10 shows the Theil estimator outperformed the MG2SLS and Fair estimators.

Case 3: In this case X_5 and X_6 are stochastic non-trended as in the first experiment and

$X_1 = 0.6\ X_{1,-1}$

$X_2 = 0.7\ X_{2,-1}$

$X_3 = 0.8\ X_{3,-1}$

$X_4 = 0.9\ X_{4,-1}$

Table 5-11 shows that the Theil's estimator which employs all T observations is performing only slightly better than its analogue MG2SLS which omits the first observation. Also as was observed in case one, the Theil estimator performed better than the Fair estimator, possibly due to the presence of high degree of trend in the explanatory variables that appear in the first equation.

To summarize, the above results show that when the explanatory variables were trended, the Theil estimator which uses all T observations, always performed better than its analogue, MG2SLS, which utilizes T-1 observations. These results also showed that whenever the exogenous variables included in the structural equation under consideration were

TABLE 5-10: Root Mean Square Error of Different Estimators Using Non-Stochastic Upward Trended Data
(r=0.6, T=30)

Coefficient	Theil	Fair	MG2SLS
C	12.65	34.70	14.07
B_1	0.24	0.42	0.23
B_2	0.34	0.51	0.34
C_1	0.21	0.39	0.22
C_2	0.40	0.75	0.40
r	0.40	0.47	0.41
Trace(MSE)*	0.54	1.37	0.55

*This index does not include the intercept term.

TABLE 5-11: Root Mean Square Error of Different Estimators Using Non-Stochastic Trended Data
(r=0.6, T=30)

Coefficient	Theil	Fair	MG2SLS
C	20.45	25.26	21.64
B_1	0.38	0.51	0.38
B_2	0.38	0.51	0.39
C_1	0.36	0.50	0.45
C_2	0.61	0.83	0.65
r	0.56	0.70	0.61
Trace(MSE)*	1.10	1.95	1.29

*This index does not include the intercept term.

strongly trended while the other X's were stochastic non-trended, Fair estimator did poorly relative to the Theil estimator. The performance of Fair relative to the Theil estimator improved whenever the degree of trend in the exogenous variables outside the structural equation under consideration increased. These results can be explained by recalling that the Fair estimator uses as instruments only the X's that appear in the structural equation under consideration. On the other hand, the Theil estimator uses as instrument all the X's included in the system . Therefore, the existence of trend in the exogenous variables outside the structural equation under consideration does not directly affect the performance of the Fair estimator. On the other hand, the Fair estimator is directly affected whenever the only trended X's are the ones included in the structural equation under consideration.

Perhaps a more important observation is that the above results seem to suggest that the effect of the trended data on the performance of the simultaneous equation estimators is different from that on the single equation estimators. Appendix III, further investigates this important observation. In Appendix III, the findings of Taylor(1981) for the single equation models are reviewed. Then, the observed differences between the simultaneous equation and the single equation models are explained. It is shown that the results of the single equation methods, regarding the importance of the first observation when the explanatory

variables are trended, are derived on the basis of very restrictive models. Finally, the investigation in Appendix III clarifies our findings in the simultaneous equation system that there does not exist a dramatic difference between the limited information methods which employ T or T-1 observations when the exogenous variables are trended. The findings of Doran and Griffitts(1982) also support our results. They conducted experiments using the seemingly unrelated regression model with first order autoregressive disturbances. Their experiment was similar to that of Maeshiro's(1980). They were particularly interested in finding the relative efficiency of the estimators which employ the first observation with those that omit it. They concluded that "in general, we feel that Maeshiro(1980) has overstated the case for inclusion of the initial transformed observation...".[47] Therefore, following conventional Monte Carlo studies, in the rest of the experiments discussed in the next section, we shall only employ stochastic non-trended exogenous variables.

B. General Results of the Monte Carlo Experiments

Table 5-12 shows the structure of the remaining experiments conducted in this part of the study. The seven estimators employed in this part are: Fair, Theil G2SLS (Theil(PW)), MG2SLS, ALIML, GLIML, 2SLS and Theil(True).[48]

---

[47]Doran, H.E., and W.E. Griffitts(1982), p.26.
[48]Note that from the family of Theil type estimators we only included the one which performed best, even though the other two excluded versions performed better than some of the

TABLE 5-12: Structure of the Monte Carlo Experiments

Disturbances	Observations	Data	Estimators
r =0.2, r =0.6, r =0.9	30, 60	Stochastic Non-Trended	Fair
			MG2SLS
r =0.9, r =0.6, r =0.2	30, 60	Stochastic Non-Trended	ALIML
			GLIML
r =0.6, r =0.9, r =0.2	60	Stochastic Non-Trended	Theil(P.W)
			Theil(True)
			2SLS

The 2SLS method estimates the structural coefficients with the restriction that the autocorrelation coefficient is equal to zero. This is, in fact, a misspecification which takes account of the simultaneity but ignores the autocorrelation. The Theil(True) method estimates the structural coefficients under the assumption of known autocorrelation coefficient.

In what follows, we shall compare the estimators according to their aggregate bias as measured by the Euclidian distance, normalized det(MSE) and normalized trace(MSE).[49] The detail statistics used to construct these indices are given in Appendix IV.

1. **Comparison using bias as a criterion of evaluation**

Table 5-13 presents the results of the antithetic estimates of the aggregate bias of different estimators when sample size is 30 and 60. For a low value of autocorrelation coefficient (r_1=0.2) and the sample size of 30, the GLIML and Theil estimators have the lowest aggregate bias, while the Fair estimator has the highest. For higher values of the autocorrelation coefficient and the sample size of 30, MG2SLS and Theil estimators have the lowest, while the Fair

---

[48](cont'd)estimators which we considered in our final experiment. This choice was made due to our interest in covering a variety of methods rather than comparing different versions of one method. Also the inclusion of the LIML estimators makes our work comparable to that of Beach and Mackinnon(1978) in the single equation context.

[49] To remind the reader, we refer to normalized det(MSE) as the index suggested by Dhrymes(1971) to signify

$$\text{Normalized det(MSE)} = \det(MSE_k)/ \det(MSE_t)$$

where MSE_t is the MSE of the Theil(True) estimator.

TABLE 5-13: General Results on the Aggregate Bias

Autocorrelation Coefficient	T	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS*
0.2	30	0.36	0.38	0.79	0.49	0.20	0.38
	60	0.10	0.10	0.71	0.20	0.04	0.08
0.6	30	0.31	0.31	0.61	0.34	0.35	0.41
	60	0.09	0.10	0.63	0.07	0.13	0.18
0.9	30	0.24	0.24	1.43	0.37	0.32	0.53
	60	0.06	0.07	1.95	0.06	0.09	0.15

*Note that the aggregate bias of 2SLS is on the basis of the four structural coefficients biases while that of other estimators include the estimated bias of the autocorrelation coefficient as well.

estimator again has the highest aggregate bias. The important observation is that the aggregate bias of 2SLS which ignores the autocorrelation, increases as the autocorrelation coefficient increases. Asymptotic theory suggest that, in the absence of lagged endogenous variables, 2SLS which takes account of simultaneity but ignores autocorrelation problem, yields consistent estimates of the structural coefficients. In other words, the consistency of 2SLS does not depend on the degree of autocorrelation. Hence it should not behave differently from other consistent estimators, in responce to the change in the degree of autocorrelation. However, the above results suggest that the small sample bias of 2SLS is positively related to the degree of autocorrelation.

As was noted before, the index of the aggregate bias does not include the constant term. In estimation of the constant term, for all degrees of autocorrelatin, the Fair estimator had the highest bias. This high degree of under-estimation of the constant term, can have serious effects on the usage of this estimator for prediction purposes.

In estimation of the autocorrelation coefficient, the Theil estimator which uses the Prais-Winsten formula had the lowest bias for all degrees of autocorrelation.

As sample size increased from 30 to 60, all estimators (except Fair) experienced reduction in their bias. In other words, increasing the sample size to 60 was enough for most

of the estimators to show their asymptotic unbiasedness. Fair's estimator again had the highest aggregate bias. The Theil estimator showed the lowest bias in estimating the autocorrelation coefficient.

2. Comparison using normalized det(MSE) criterion

A more illuminating picture of the performance of different estimators can be observed by examining Table 5-14 in which we have normalized the det(MSE) of different estimators by dividing it by its corresponding det(MSE) of the Theil(true) estimator. When sample size is equal to 30 we can observe that for low degrees of autocorrelation, i.e., r_1=0.2, 2SLS has the lowest normalized det(MSE), followed by the Theil and MG2SLS estimators. But as the autocorrelation coefficient increases, the performance of 2SLS deteriorates and it becomes inferior to the Theil estimator when the autocorrelation coefficient becomes very high, i.e.,r_1=0.9. This behaviour of the 2SLS confirms the asymptotic theory that suggest a serious loss of efficiency for those estimators which ignore the autocorrelation problem.

Comparison of the Theil and MG2SLS estimators reveals that the Theil estimator which utilizes all T observations has lower det(MSE) than the MG2SLS which omits the first observation, for all degrees of autocorrelation. Therefore, utilization of the first observation seems to have positive

TABLE 5-14: General Results on the Normalized Determinant(MSE)*

Autocorrelation Coefficient	T	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
0.2	30	4.04	5.74	161.81	3425781	731.17	1.49
	60	1.26	1.58	4463	10031	40.42	1.20
0.6	30	1.02	4.01	39.07	11669	275.84	2.23
	60	1.39	1.58	33804	11.48	79.35	136.96
0.9	30	2.13	2.42	14483	55681	277.25	154.13
	60	0.97	1.03	159190	17.05	12.77	46184.9

*Note this index does not include the intercept.

effect on the performance of the Theil estimator.[50] Comparison of GLIML and ALIML revealed that the GLIML estimator outperformed ALIML for all degrees of autocorrelation when the sample size was 30.

Table 5-14 also demonstrates that the above picture stays more or less the same as the sample size increases from 30 to 60 observations. For a very low degree of autocorrelation, 2SLS again performed best, but its performance deteriorated rapidly as the autocorrelation coefficient increases. It became extremely inefficient when the degree of autocorrelation reached 0.9. The performance of ALIML and GLIML improved considerably (especially for r=0.6 and r=0.9) as the sample size increased. Finally, the efficiency gain caused by employment of the first observation does not disappear even when the sample size was increased to 60. The Theil estimator which utilizes all T observations is still superior to the MG2SLS estimator, for all degrees of autocorrelation. This fact suggests that employment of the first observation does increase the efficiency of estimators even for moderately large sample sizes.

---

[50] The fact that some of the entries in Tables 5-14 and 5-15 are equal or less than one, suggest equivalence or higher efficiency of a particular estimator relative to Theil(true) in estimating some of the coefficients.

3. Comparison using normalized trace(MSE) criterion

A slightly different picture appears when we compare estimators on the basis of the normalized trace(MSE). It should be mentioned that, as was the case with normalized det(MSE), the normalized trace index does not include the constant term. We omitted the intercept term since the complete trace was often dominated by the MSE of the intercept term. Since researchers are often interested in obtaining reliable estimate of the slope coefficients, the complete trace index could provide misleading assessment.

Comparing estimators on the basis of the normalized trace, we can see (Table 5-15) that for sample size of 30, 2SLS is no longer best even for low degrees of autocorrelation, i.e., r=0.2. The relative performance of 2SLS deteriorates rapidly as the autocorrelation coefficient increased. This picture stayed unchanged as we increased the sample size to 60. In most cases, except for sample size of 30 and r of 0.9, the Theil estimator performed marginally better than MG2SLS. This suggests a gain in efficiency due to the employment of the first observation even for a relatively large sample size, i.e., n=60. Another important observation is that, except for one case(r=0.6, N=60), the relative performance of the Fair estimator deteriorated as the autocorrelation coefficient or the sample size increased. This obviously suggests that the gain in efficiency achieved by Theil(True) estimator, which is used as a base, has been much greater than that of the Fair

TABLE 5-15: General Results on the Normalized Trace(MSE)*

Autocorrelation Coefficient	T	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
0.2	30	1.14	1.24	3.60	244.99	3.44	1.20
	60	1.02	1.08	22.49	34.55	1.08	1.15
0.6	30	1.02	1.21	3.57	16.10	3.67	1.53
	60	1.01	1.14	36.90	1.85	2.13	4.36
0.9	30	1.64	1.30	25.64	14.35	6.96	5.42
	60	1.00	1.12	347.49	2.50	1.63	19.22

*Note this index does not include the intercept.

estimator when the sample size was doubled. This observation cast some doubt on the usefulness of the Fair estimator for estimation of models characterized by autocorrelation, at least for the case of stochastic non-trended data.

To summarize, the above results provide important information in guiding a potential econometric practitioner in the choice of estimator for models characterized by autocorrelation. Theil's estimator generally dominated all other methods. Employment of the first observation improved the performance of the Theil estimator. However, the efficiency gain was not as large as was suggested by Park and Mitchell(1980) and Maeshiro(1979). This results are in line with Doran and Griffiths'(1982) findings in the single equation context. Generalized LIML, even though inferior to Theil estimator, dominated ALIML in most cases which suggests that imposing the restriction that all autocorrelation coefficients are equal, even when this is not true, does not impose very large costs in terms of MSE. The Fair estimator which is a commonly used method for estimation of the autocorrelated models performed poorly. In most cases it proved to be inferior even to 2SLS which ignores the autocorrelation.

C. The Performance of Estimators in Hypothesis Testing

Performance of different estimators in test of hypothesis is as important for an econometric practitioner as obtaining efficient estimates of the structural

coefficients.

There are number of tests of hypotheses proposed in the literature. Maddala(1974) using a two equation model, compared the power functions of the tests of significance proposed by Dhrymes(1969), Richardson and Rohr(1971) and Anderson and Rubin(1949) with that of the conventional tests of significance based on asymptotic theory. He found that the conventional test performed better than the alternatives proposed.

However, it must be noted that the comparison of the performance of different estimators in hypothesis testing has not received enough attention. Park and Mitchell(1980) using the conventional tests of significance, examined the performance of different single equation estimators when the model possessed autocorrelation. They found that none of the single equation estimators they considered performed well in hypothesis testing. They found that the percentage of times the true null hypothesis was rejected was much greater than the nominal level of significance. They attributed this problem to the underestimation of the covariance matrix even when the Prais-Winsten estimator, which they found to give a accurate estimate of the standard error, was used . They considered this problem as serious as obtaining inefficient estimates of the structural coefficient.

Following Park and Mitchell, we focus on the number of times each method leads to the occurrence of a Type I error at the 5 pecent level of significance. Detailed statistics

concerning the Type I error and the power of the test statistics for different sample sizes and different degrees of autocorrelation are given in the Appendix VI. In this Chapter we shall only provide aggregate measures; namely, the average Type I error and the average power of the test statistics accross equations. However, we shall accompany these averages with the statistics concerning the range of the Type I error and the power of tests across coefficients. Table 5-16 shows the comparison of the estimators in hypothesis testing for different values of autocorrelation coefficient and sample sizes of 30 and 60. The figures in parantheses are the range of the Type I errors. For a sample size of 30 we can observe that for low degree of autocorrelation, i.e., r= 0.2, GLIML performed best and was followed by 2SLS, Theil and MG2SLS estimators. But as the autocorrelation coefficient increased, Theil and its modified version outperformed all others. In fact, their observed Type I error moved toward the nominal level of significance as the autocorrelation coefficient increased. The reverse of this process occurred in the case of Fair's estimator. Its observed Type I error increased as the autocorrelation coefficient increased from 0.2 to 0.9.

Generally, for sample size of 30 the observed Type I errors of all estimators (except for Theil when r=0.9) was much greater than the nominal level of significance, suggesting that the tails of the test statistic distribution were much thicker than those given by the t-distribution.

TABLE 5-16: Average Type I Error at 5% Significant Level

Autocorrelation Coefficient	T	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
0.2	30	27 (13-32)	26 (18-31)	35 (6-66)	33 (18-46)	19 (11-25)	27 (16-33)
	60	12 (9-16)	13 (10-17)	47 (30-54)	10 (5-16)	10 (8-13)	12 (9-15)
0.6	30	18 (10-22)	17 (11-22)	35 (10-54)	26 (12-40)	20 (12-26)	25 (7-36)
	60	8 (4-12)	8 (3-11)	41 (21-49)	6 (3-10)	14 (6-16)	18 (8-23)
0.9	30	9 (6-13)	11 (7-14)	40 (19-59)	27 (14-37)	14 (8-19)	26 (9-36)
	60	6 (1-8)	6 (1-8)	90 (53-99)	7 (3-9)	8 (4-10)	27 (9-34)

This result is in line with the conclusions reached by Park and Mitchell in the single equation case.

As the sample size increased to 60 (Table 5-16), all estimators (except Fair) experienced a significant reduction in the actual Type I error. In fact, their observed Type I error became closer to the nominal level of significance. For the low autocorrelation coefficient of 0.2, all estimators (except Fair) performed more or less the same. The increased sample size had a positive effect on the performance of ALIML. In fact, ALIML had the lowest Type I error for the sample size of 60 and the autocorrelation coefficients of 0.2 and 0.6. The performance of 2SLS and Fair estimators deteriorated as the autocorrelation coefficient increased to 0.9. These results suggest that perhaps the main problem with the application of 2SLS and Fair estimators to the autocorrelated models is that they result in much greater probability of Type I error than the nominal level of significance indicates.

Table 5-17 presents the power of the test of significance at 5 percent significant level when the sample size is equal to 30 and 60. Again the figures in parantheses show the range of the power of the test across coefficients. For a sample size of 30 we can observe that the ALIML showed the highest power for all degrees of autocorrelation. The power of the 2SLS estimator deteriorated as the autocorrelation coefficient increased, while that of the Theil, MG2SLS, Fair and GLIML remained more or less the

TABLE 5-17: Average Power of Test of Significance at 5% Significant Level

Autocorrelation Coefficient	T	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
0.2	30	40 (1-90)	42 (1-92)	4 (4-95)	54 (29-84)	40 (6-83)	41 (1-91)
	60	62 (11-10)	63 (11-100)	39 (3-89)	78 (35-100)	68 (15-100)	61 (11-98)
0.6	30	39 (2-85)	40 (3-87)	44 (6-92)	55 (27-87)	41 (6-86)	37 (1-91)
	60	66 (14-99)	65 (10-99)	42 (8-92)	85 (52-100)	65 (16-99)	50 (7-95)
0.9	30	42 (4-91)	43 (4-91)	43 (12-93)	58 (41-85)	42 (6-85)	33 (3-89)
	60	72 (23-99)	73 (23-99)	61 (6-100)	85 (51-99)	73 (25-100)	35 (3-85)

same. The relatively high power of the ALIML, Fair and GLIML estimators is misleading. The relatively high power of these estimator can be attributed to their underestimation of the standard error that in turn leads to the high probability of Type I and low probability of Type II errors. It is conceivable that if the nominal level of significance be lowered to the point where the observed Type I error of these estimators be equal to that of the Theil estimator at 5 percent level, their power will be much lower than the above picture suggests.

As the sample size increased to 60 (Table 5-17), the observed power of all estimators (except Fair) increased. The power of 2SLS decreased as the autocorrelation coefficient increased. ALIML showed the highest power.

To summarize, the results of this Chapter suggest important guidelines for econometric practitioners. In practice a researcher usually deals with a small or moderate sample sizes. For the sample sizes of about 30 observations and when the autocorrelation coefficient is very low, i.e., r=0.2, 2SLS estimator which ignores the autocorrelation performs as well as any other estimator which corrects for the presence of serial correlation in the disturbance term. As the autocorrelation coefficient rises, the performance of 2SLS deteriorates rapidly. Fair's estimator which is probably the most commonly known method did not perform much better than the 2SLS. In fact, in most cases it was inferior to the 2SLS which ignores the autocorrelation. The Theil

estimator which is completely ignored in empirical research turns out to be the best not only in the estimation of the structural coefficients, but also in the test of hypothesis.

For the larger sample sizes of about 60, the Theil estimator performed best in the estimation of the structural coefficients but was outperformed by ALIML in the hypothesis testing.

VI. LIMITED INFORMATION METHODS OF ESTIMATION IN THE PRESENCE OF LAGGED ENDOGENOUS VARIABLES

A. Statement of the Problem

Estimation problems associated with the presence of autocorrelation in simultaneous equation systems in the absence of lagged endogenous variables were discussed in Chapters I and II. In this chapter we will consider problems of estimation related to the presence of lagged endogenous variables in simultaneous equation systems. We shall also analyse different methods appropriate for estimation of this type of system.

The presence of lagged endogenous variables among the explanatory variables of a simultaneous equation system with autocorrelated errors creates additional estimation problems. Consider the following system

$$Y B + X^* \Gamma = U \qquad (1)$$

$$U = U_{-1} R + E$$

where Y is the T.G matrix of observations on the endogenous variables. X* is the T.K matrix of observations on the predetermined variables, i.e.,$X^* = (Y_{-1}, X)$, where Y_{-1} is a matrix of endogenous variables lagged one period, and X is the matrix of purely exogenous variables of the system. B is the G.G matrix of structural coefficients associated with the current endogenous variables. Γ is the K.G matrix of coefficients of the predetermined variables. U is the T.G matrix of disturbances.

The first structural equation of this system can be written as

$$y_1 = Y_1\beta_1 + X_1{*}\,\gamma_1 + u_1 \tag{2}$$

$$u_1 = r\,u_{1,-1} + \epsilon_1$$

$$X_1{*} = (Y_{1,-1}\ ,\ X_1)$$

where y_1 and Y_1 are appropriate sub-matrices of Y; X_1, r, u_1 and ϵ_1 are appropriate sub-matrices of X, R, U, and E, respectively. $Y_{1,-1}$ is a subset of Y, lagged one period, and X_1 is a sub-matrix of X.

In the absence of lagged endogenous variables, problems of estimation of system (2) are the correlation between Y_1 and u_1, and the autocorrelation of the error term, u_1. The presence of lagged endogenous variables in a simultaneous equation system without autocorrelation does not create any additional estimation problems. The only problem would be the correlation between the current endogenous variables appearing among the explanatory variables of the system and the error term. There will not be any correlation of the lagged endogenous variables and the current error terms. However, when the current error terms are correlated with their past values, they will be also correlated with the lagged endogenous variables. Therefore, the presence of the lagged endogenous variables poses additional estimation problems when a model is characterized by autocorrelation.

In the absence of lagged endogenous variables, application of standard techniques which ignored the autocorrelation property of the error terms resulted in

consistent but inefficient estimates of the structural coefficients. However, when the structural equation includes lagged endogenous variables, standard techniques no longer provide consistent estimates of the structural coefficients. This is due to the fact that in this case the predetermined variables are also correlated with the error terms. Therefore, estimators appropriate for estimation of autocorrelated models with lagged endogenous variables are those which take into account both the simultaneity and the correlation of the predetermined variables with the error terms.

B. Methods of Estimation

There are a number of estimators designed for estimation of the model (2). Two of these estimators, namely Fair and modified Brundy and Jorgenson were discussed in Chapter II. The structure of these estimators remains the same except that in this case the X matrix is not purely exogenous, but contains lagged endogenous variables. Asymptotic properties of these two estimators are also the same as they were in the case with no lagged endogenous variable.

Theil Estimator

A limited information method that can be employed for estimation of the model (2) is a modified version of Theil's

generalized two stage least squared estimator (G2SLSM).[51] A modification is needed since the procedure outlined in Chapter II for the G2SLS estimator is not applicable to models with lagged endogenous variables. The reason is that in the absence of lagged endogenous variables, a consistent estimate of the autocorrelation coefficient of the structural equation under consideration was used to construct the Prais-Winsten transformation matrix. This matrix was then employed to transform the ordinary reduced form of the system as

$$P\ Y = P\ X\ \Pi_1 + P\ U\ B^{-1} \tag{3}$$

A consistent estimate of PY could then be obtained by application of OLS to (3). However, if X includes lagged endogenous variables, the transformed X would be correlated with the error term PUB^{-1} which includes u_i and $u_{i,-1}$ for $i=1,...,G$.[52] Therefore, a consistent estimate of PY cannot be obtained through this approach. To remedy this problem we propose the following two alternatives:

1- Write the transformed ordinary reduced form of the system (1) as

----

[51]Note that G2SLSM is different from the modified G2SLS(MG2SLS) introduced in Chapter II. MG2SLS was similar to G2SLS except for the omission of the first observation. The G2SLSM estimator discuused in this chapter is the analogue of G2SLS which takes account of lagged endogenous variables present in the structural equation under consideration.

[52]Note that The transformed X's will be uncorrelated with the error terms only if all the autocorrelation coefficients of the structural equations are equal, i.e., $PUB^{-1}=EB^{-1}$.

$$P\ Y = P\ Y_{-1}\Pi_1* + P\ X\ \Pi_2* + P\ U\ B^{-1}$$

or

$$\dot{Y} = \dot{Y}_{-1}\Pi_1* + \dot{X}\ \Pi_2* + V \qquad (4)$$

Obtain a consistent estimate of PY by applying the instrumental variable method to (4), using a proper instrument for Y_{-1}, i.e., $\hat{Y}_{-1}=X_{-1}\hat{\Pi}$. If W_1 is the set of instruments so selected, we will have

$$\hat{\Pi}_{IV} = (\dot{W}'\dot{X}*)^{-1}\dot{W}'\dot{Y}$$

where

$$\dot{W} = (PW_1 \quad PX)$$

$$\dot{X}* = (PY_{-1} \quad PX)$$

$$\hat{\Pi}_{IV} = (\hat{\Pi}_1*' \ \hat{\Pi}_2*')'$$

therefore

$$\hat{\dot{Y}} = \dot{Y}_{-1}\ \hat{\Pi}_1* + \dot{X}\ \hat{\Pi}_2* \qquad (5)$$

2-Alternatively, we can obtain a consistent estimate of PY by employing the augmented reduced form of the system (1)

$$Y = Y_{-1}\Pi_1+Y_{-2}\Pi_2+X\Pi_3+X_{-1}\Pi_4+EB^{-1} \qquad (6)$$

A consistent estimate of PY can be obtained by transforming (6) and estimating the resultant transformed equation by OLS. Having estimated PY consistently, Theil's G2SLSM estimator of the system (2) can be obtained as follows:

Write the model presented in (2) as

$$y_1 = Y_1\beta_1+Y_{1,-1}\gamma_{11}+X_1\gamma_{12}+u_1 \qquad (7)$$

If the autocorrelation coefficient is known to be equal to r, we can use the Prais-Winsten transformation on (7) to obtain[53]

----

[53] Note that we shall use the Prais-Winsten transformation matrix only if a consistent estimate of PY is obtained

$$Py_1 = PY_1\beta_1 + PY_{1,-1}\gamma_{11} + PX_1\gamma_{12} + Pu_1$$

or

$$\dot{y}_1 = \dot{Y}_1\beta_1 + \dot{Y}_{1,-1}\gamma_{11} + \dot{X}_1\gamma_{12} + \epsilon_1 \qquad (8)$$

Theil's G2SLSM estimator of the coefficients of the system (7) can be obtained by replacing $\dot{Y}_1$ in (8) by its consistent estimate obtained above and estimating the resultant equation by OLS; that is,

$$\hat{\delta}_1 = (\hat{\dot{Z}}_1'\hat{\dot{Z}}_1)^{-1}\hat{\dot{Z}}_1'\dot{y}_1 \qquad (9)$$

where

$$\hat{\dot{Z}}_1 = (\hat{\dot{Y}}_1\ \dot{Y}_{1,-1}\ \dot{X}_1)$$

$$\delta_1 = (\beta_1'\ \gamma_{11}'\ \gamma_{12}')'$$

Under the assumption of known and equal autocorrelation coefficients, the G2SLSM estimator (9) is equivalent to a usual 2SLS estimator of the transformed system (8) and therefore is consistent and efficient within the class of limited information estimators. Its asymptotoic covariance matrix is

$$\text{Asy-cov}(\hat{\delta}_1) = \sigma_{11}\text{plim } T(\hat{\dot{Z}}_1'\hat{\dot{Z}}_1)^{-1} \qquad (10)$$

However, new problems of estimation arise when the autocorrelation coefficient is unknown and has to be replaced by its consistent estimate. A consistent estimate of the autocorrelation coefficient can be obtained from the estimated residuals of the system (7). To estimate (7), we not only have to consider the correlation between Y_1 and u_1,

----

[53](cont'd)through the estimation of the ordinary reduced form. If the augmented reduced form is used and therefore only T-1 fitted values for PY are obtained, we will use the CORC transformation matrix which disregards the transformation of the first observation.

but also should take into account of the correlation between $Y_{1,-1}$ and u_1. This means that at the initial stage we should use an instrumental variable estimator which replaces Y_1 and $Y_{1,-1}$ by proper instruments, possibly from the list of the current and lagged values of the exogenous variables excluded from the structural equation (7). In other words, if W is the matrix of instruments so selected, the first stage instrumental variable estimator of δ_1 is

$$\hat{\delta}_{IV} = (W'Z_1)^{-1}W'y_1 \tag{11}$$

where

$$Z_1 = (Y_1 \; Y_{1,-1} \; X_1)$$

Under standard assumptions the estimator in (11) is consistent. Then $\hat{\delta}_{IV}$ can be used to obtain a consistent estimate of the residuals u_1, and thus calculate a consistent estimate of the autocorrelation coefficient, denoted as $\hat{r}$. Using $\hat{r}$ we can transform the structural equation (7) and estimate the structural coefficient following the procedure outlined above.

Following the approach adopted by Fair(1970), we can show the consistency of Theil's G2SLSM estimator when it uses a consistent estimate of the autocorrelation coefficient. To do this we can write equation (8) as[54]

$$\begin{aligned} \bar{y}_1 = \hat{\bar{Y}}_1\beta_1 + \bar{Y}_{1,-1}\gamma_{11} + \bar{X}_1\gamma_{12} \\ + [u_{1,-1}(r-r) + \epsilon_1 + \hat{\bar{V}}_1\beta_1] \end{aligned} \tag{12}$$

where

----

[54] Note that for simplicity we shall ignore the transformed first observation which is asymptotically negligible.

$$\hat{\bar{V}}_1 = \bar{Y}_1 - \hat{\bar{Y}}_1$$

and $\hat{\bar{Y}}_1$ is a consistent estimate of $\bar{Y}_1$ obtained from the ordinary or augmented reduced form. Theil's G2SLSM estimator of the structural coefficients can be obtained by estimating equation (12) by OLS. Application of OLS to equation (12) yields the estimated coefficients which minimizes the sum of squared residuals. Therefore, If the three components of the error term are orthogonal to each other, the sum of squared residuals occurs where the square of its each component term is at its minimum. The square of the first term reaches its minimum when $r=\hat{r}$ which in turn eliminates $u_{1,-1}$ from the residual term and ensures the consistency. Therefore, Theil's estimator is consistent only if $u_{1,-1}$ and $\hat{\bar{V}}_1$ are orthogonal to each other. This is so, since if they are not orthogonal, the minimum sum of squared residuals of (12) will not occur where $r = \hat{r}$. This in turn leaves the $u_{1,-1}$ in the residuals term that is correlated with $\bar{Y}_{1,-1}$ and therefore renders Theil's estimator inconsistent.

In the case where the augmented reduced form is used to obtain a consistent estimate of $\bar{Y}$, $u_{1,-1}$ and $\hat{\bar{V}}_1$ are orthogonal since $u_{1,-1}$ is a function of Y_{-1}, Y_{-2} and X_{-1} (see 14 below) and these variables are used as instruments in the estimation of $\hat{\bar{V}}_1$. Therefore, Theil's G2SLSM estimator will be consistent. Proof of the consistency of G2SLSM when it uses the ordinary reduced form also requires the orthogonality of $u_{1,-1}$ and $\hat{\bar{V}}_1$. To demonstrate the orthogonality of $u_{1,-1}$ and $\hat{\bar{V}}_1$ when the consistent estimate

of $\overline{Y}$ is obtained through the ordinary reduced form equation (4), write the system (1) as

$$Y\,B + Y_{-1}\Gamma_1 + X\,\Gamma_2 = U \tag{13}$$

From (13), we have

$$\begin{aligned} U_{-1} &= Y_{-1}B + Y_{-2}\Gamma_1 + X_{-1}\Gamma_2 \\ &= (Y_{-1}\ Y_{-2}\ X_{-1})\Xi \end{aligned} \tag{14}$$

Also equation (5) can be written as

$$\hat{\overline{Y}} = (Y_{-1}\ Y_{-2}\ X\ X_{-1}) \begin{pmatrix} I & 0 \\ -rI & 0 \\ 0 & I \\ 0 & -rI \end{pmatrix} \begin{pmatrix} \hat{\Pi}_1* \\ \hat{\Pi}_2* \end{pmatrix} \tag{15}$$

Equation (14) shows that U_{-1} is a function of Y_{-1}, Y_{-2} and X_{-1}. Equation (15) demonstrates that these variables are also used as instruments in obtaining $\hat{\overline{Y}}_1$ and thus in calculating $\hat{V}_1$. Therefore, by the property of OLS, $\hat{V}_1$ and U_{-1} are orthogonal. Hence, Theil's G2SLSM estimator when it utilizes a consistent estimate of the autocorrelation coefficient is consistent.

Dhrymes' Estimators

Dhrymes,et al.(1974) proposed a number of alternative estimators for estimation of the model (2). These estimators can be reduced to two basically different methods. One method is a two-step instrumental variable estimator and employs the ordinary reduced form to obtain predictions of the current and lagged endogenous variables appearing among the explanatory variables of the structural equation under

consideration. The second method which they called the "converging iterate two stage least squares autoregressive estimator (C2SLSA)" is a generalization of the Fair estimator discussed in Chapter II.

The two-step instrumental variable estimator designed for estimation of system (2) is constructed as follows:

1- Write the system under consideration as

$$Y\ B + Y_{-1}\Gamma_1 + X\ \Gamma_2 = U \tag{16}$$

Estimate each structural equation by an instrumental variable estimator and form a consistent estimate of B, Γ_1, and Γ_2. Using these consistent estimates, form the reduced form of the system (22) and obtain consistent predictions of the current dependent variable as

$$\hat{Y}_1 = Y_{-1}\hat{\Pi}_1 + X\ \hat{\Pi}_2 \tag{17}$$

where

$$\hat{\Pi}_1 = -\hat{\Gamma}_1\hat{B}^{-1}$$

$$\hat{\Pi}_2 = -\hat{\Gamma}_2\hat{B}^{-1}$$

For (17) to be operational, we need a starting value for Y at time zero, i.e., Y_0. It is permissable to choose an initial condition of Y_0=0. Therefore, using (17) and the starting value for Y, we obtain fitted values of $\hat{Y}_1$ and their one period lags, $\hat{Y}_{1,-1}$.

2-Obtain a consistent estimate of the autocorrelation coefficient of the structural equation under consideration, using a consistent estimate of the residuals calculated from the instrumental variable estimation at the first step. Define $\dot{W}_1$, $\dot{Z}_1$,$\dot{y}_1$ and δ_1

$$\dot{W}_1 = [P\hat{Y}_1 \ \ P\hat{Y}_{1,-1} \ \ PX_1]$$

$$\dot{Z}_1 = [PY_1 \ \ PY_{1,-1} \ \ PX_1]$$

$$\dot{y}_1 = Py_1$$

$$\delta_1 = (\beta_1' \ \ \gamma_{11}' \ \ \gamma_{12}')'$$

Where P is the Prais-Winsten transformation matrix. Then, the proposed two-step instrumental variable estimator is

$$\hat{\delta}_1 = (\dot{W}_1'\dot{Z}_1)^{-1}\dot{W}_1'\dot{y}_1 \tag{18}$$

This two-step estimator is very closely related to the instrumental variable method proposed by Brundy and Jorgenson(1971) for systems with no autocorrelation. It can also be seen that this two-step estimator is nothing more than a method which treats the lagged endogenous variable as an endogenous variable and replaces it with its systematic part in the estimation.

The second method, C2SLSA, is constructed as follows:

1- Obtain the augmented reduced form of the system (1) as

$$Y = Y_{-1}\Pi_1+Y_{-2}\Pi_2+X\ \Pi_3+X_{-1}\Pi_4+w \tag{19}$$

Regress each column of Y_1 on Y_{-1}, Y_{-2}, X and X_{-1} and obtain the predicted values of Y_1, denoted as $\hat{Y}_1$.[55]

2- Estimate the structural equation (2) with an instrumental variable estimator and obtain a consistent estimate of the autocorrelation coefficient r using the estimated residuals calculated from the instrumental variable estimation.

---

[55] Note that Y_1 can also be calculated from (19) by replacing Π_i, $i=1,\ldots,4$, by their consistent estimates.

3- Using the estimated r, transform the structural equation (2) as

$$y_1-\hat{r}y_{1,-1} = (Y_1-\hat{r}Y_{1,-1})\beta_1+(X_1*-\hat{r}X_{1,-1}*)\gamma_1+\epsilon_1$$

or

$$y_1-\hat{r}y_{1,-1} = (Y_1-\hat{r}Y_{1,-1})\beta_1 +[(Y_{1,-1}-\hat{r}Y_{1,-2}),(X_1-\hat{r}X_{1,-1}]\gamma_1+\epsilon_1 \quad (20)$$

4- Define $\hat{\bar{Z}}_1$, $\bar{y}_1$, and δ_1 as

$$\hat{\bar{Z}}_1 = [(\hat{Y}_1-\hat{r}Y_{1,-1}),(Y_{1,-1}-\hat{r}Y_{1,-2}),(X_1-\hat{r}X_{1,-1})]$$

$$\bar{y}_1 = y_1 - \hat{r}y_{1,-1}$$

$$\delta_1 = (\beta_1' \quad \gamma_1')'$$

Estimate the structural coefficient as

$$\hat{\delta}_1 = (\hat{\bar{Z}}_1'\hat{\bar{Z}}_1)^{-1}\hat{\bar{Z}}_1'\bar{y}_1 \quad (21)$$

5- Using $\hat{\delta}_1$, recompute r from step (3) and continue until convergence.

Dhrymes, et al. showed that both the two-step instrumental variable and C2SLSA estimators are consistent. They also showed that the C2SLSA estimator that uses the augmented reduced form to obtain predictions of the current dependent variables is more efficient than the instrumental variable method which uses the ordinary reduced form. Under the assumption of known coefficient of autocorrelation, the asymptotic covariance matrix of the C2SLSA estimator is equal to

$$\text{Asy-cov}(\hat{\delta}_1) = \sigma_{11}\text{plim } T(\hat{\bar{Z}}_1'\hat{\bar{Z}}_1)^{-1} \quad (22)$$

However, when the coefficient of autocorrelation is unknown and is replaced by its consistent estimate, the C2SLSA estimator becomes asymptotically less efficient. Derivation

of the asymptotic covariance matrix of this estimator when coefficient of autocorrelation is unknown is given in Dhrymes, et al.(1974).

Hatanaka's Estimators

Hatanaka(1976) developed three alternative limited information two-step estimators for estimation of the system (2). These efficient methods are based on the technique which he developed for estimation of a dynamic single equation model with autoregressive errors (Hatanaka,1974). He called it "the Residual-Adjusted Aitken estimator" since it was basically an Aitken type estimator except that it included the estimated residuals, lagged one period, in the list of regressors to ensure asymptotic efficiency. Hatanaka's Residual Adjusted Aitken estimator is similar to a modified method of scoring for maximizing the log-likelihood function except that in its derivation the terms with zero probability limits were ignored. For finite sample sizes the Residual Adjusted Aitken estimator is not identical to the modified method of scoring.[56] However, this estimator is identical to a two-step Gauss-Newton estimator. Only one iteration of Gauss-Newton, that is a regression of the residuals, E_t, on $\partial E_t/\partial(\delta)$, yields an asymptotically efficient estimator of the structural coefficients.[57]

---

[56] Hatanaka (1974), p.202.

[57] For discussion of different methods of numerical optimization and the equivalence of Hatanaka's estimator and the two-step Gauss-Newton estimator see Harvey(1981), Chapter 4 and 8.

The three alternative limited information estimators which Hatanaka proposed using the method developed in the Residual Adjusted Aitken estimator share the following first step in common:

1-Estimate the structural equation by an instrumental variable estimator and obtain a consistent estimate of the residuals, u_1. Then use the estimated residuals to obtain a consistent estimate of the autocorrelation coefficient r, denoted as r, using the Cochrane-Orcutt method.

The second step for these three alternative methods denoted as A, B, and C, is as follows

Method A:

Use the augmented reduced form

$$Y = Y_{-1}\Pi_1 + Y_{-2}\Pi_2 + X\,\Pi_3 + X_{-1}\Pi_4 + w \qquad (23)$$

and regress each column of Y_1 on Y_{-1}, Y_{-2}, X, X_{-1} to obtain fitted values for Y_1, denoted as $\hat{Y}_1$. Define $\hat{\bar{Z}}_1$, $\bar{y}_1$, and δ_1 as

$$\hat{\bar{Z}}_1 = [(\hat{Y}_1 - \hat{r}Y_{1,-1}),(Y_{1,-1} - \hat{r}Y_{1,-2}),(X_1 - \hat{r}X_{1,-1}),\hat{u}_{1,-1}]$$

$$\bar{y}_1 = y_1 - \hat{r}y_{1,-1}$$

$$\delta_1 = (\beta_1{}' \quad \gamma_1{}')'$$

Note that the matrix $\bar{Z}_1$ has been augmented by the vector of lagged residuals $\hat{u}_{1,-1}$. Then estimate δ_1 and r_{11} as

$$\begin{pmatrix} \hat{\delta}_1 \\ \\ \hat{r}_{11} \end{pmatrix} = (\hat{\bar{Z}}_1{}'\hat{\bar{Z}}_1)^{-1}\hat{\bar{Z}}_1{}'\bar{y}_1 \qquad (25)$$

Finally, the estimator is defined as

$$\begin{pmatrix} \hat{\delta}_1 \\ \\ \hat{r}_1 \end{pmatrix} = \begin{pmatrix} \hat{\delta}_1 \\ \\ \hat{r} + \hat{r}_{11} \end{pmatrix}$$

This estimator can be regarded as a simplification and extension of the Sargan(1961) and Amemiya(1966) methods.

Method B:

The second estimator is a simplified form of the Brundy and Jorgenson instrumental variable method which was modified by Fair(1972) to take into account the autocorrelation property of the disturbances. The procedure followed in this method is as follows:

1- Obtain fitted values of Y_1 from the reduced form

$$\hat{Y} = Y_{-1}\hat{\Pi}_1 + Y_{-2}\hat{\Pi}_2 + X\hat{\Pi}_3 + X_{-1}\hat{\Pi}_4 \quad (26)$$

where $\hat{\Pi}_1$, $\hat{\Pi}_2$, $\hat{\Pi}_3$ and $\hat{\Pi}_4$ are consistent estimate of the reduced form coefficients formed from consistent estimates of the structural coefficient obtained from the first step Instrumental variable estimation applied to all the equations. Define $\hat{\overline{W}}_1$ and $\overline{Z}_1$ as

$$\hat{\overline{W}}_1 = [(\hat{Y}_1 - \hat{r}Y_{1,-1}), (Y_{1,-1} - \hat{r}Y_{1,-2}), (X_1 - \hat{r}X_{1,-1}), \hat{u}_{1,-1}]$$

$$\overline{Z}_1 = [(Y_1 - \hat{r}Y_{1,-1}), (Y_{1,-1} - \hat{r}Y_{1,-2}), (X_1 - \hat{r}X_{1,-1}), \hat{u}_{1,-1}]$$

Then estimate the structural coefficients with the instrumental variable method, namely

$$\begin{pmatrix} \hat{\delta}_1 \\ \\ \hat{r}_{11} \end{pmatrix} = (\hat{\overline{W}}_1{}'\overline{Z}_1)^{-1}\hat{\overline{W}}_1{}'\overline{y}_1 \quad (27)$$

The estimator is then defined by

$$\begin{pmatrix} \hat{\delta}_1 \\ \\ \hat{r}_1 \end{pmatrix} = \begin{pmatrix} \hat{\delta}_1 \\ \\ \hat{r} + \hat{r}_{11} \end{pmatrix} \tag{28}$$

Method C:

To obtain the third estimator, define

$$\hat{\bar{u}}_1 = \hat{u}_1 - r\hat{u}_{1,-1} \tag{29}$$

Then estimate the structural coefficients as

$$\begin{pmatrix} \tilde{\delta}_1 \\ \\ \hat{r}_{11} \end{pmatrix} = (\hat{\bar{W}}_1{}'\hat{\bar{W}}_1)^{-1}\hat{\bar{W}}_1{}'\hat{\bar{u}}_1 \tag{30}$$

Then the third alternative estimator is

$$\begin{pmatrix} \hat{\delta}_1 \\ \\ \hat{r}_1 \end{pmatrix} = \begin{pmatrix} \tilde{\delta}_1 + \hat{\delta}_{IV} \\ \\ \hat{r} + \hat{r}_{11} \end{pmatrix} \tag{31}$$

Where $\hat{\delta}_{IV}$ is the instrumental variable estimate of the first structural equation coefficients obtained in the first step. This estimator can be interpreted as a second step in a Gauss Newton procedure.[58]

Hatanaka showed that these three two-step estimators are consistent and asymptotically efficient within the class of limited information estimators. Therefore, the choice among them should be made in regard to their small sample performance.

----

[58] See Pagan(1982).

Other Estimators

Other limited information estimators have been proposed for estimation of dynamic autoregressive simultaneous equation systems. In fact, a large number of asymptotically efficient estimators can be defined using the estimator generating equation in Hendry(1976). As a particular case Hendry and Srba(1977) proposed an instrumental variable estimator (AIV) which minimizes a quadratic form with respect to the structural coefficients.[59] Using a two equation dynamic autoregressive model they compared the small sample properties of AIV with 2SLS, OLS and a variation of OLS which corrects for the presence of autocorrelation. They found that AIV outperformed other estimators for large samples (T>55) and considerable autocorrelation (r>0.5); 2SLS was optimal for large samples and small autocorrelation coefficients; and OLS was best for small samples and low autocorrelation coefficients.[60]

Wang and Fuller (1982) also proposed a limited information method for estimation of dynamic autoregressive models. Their estimator is an extension of the two-step Gausss-Newton procedure which is similar to the Hatanaka (Method A) estimator but was independently derived in Wang and Fuller(1975).

----

[59] See Hendry and Srba (1977), p.971.

[60] Note that this estimator was not known to us at the time of planning for simulation and therefore is not used in this thesis.

C. Small Sample Investigation

To investigate the small sample properties of the limited information estimators designed for the estimation of the dynamic simultaneous equations models with autocorrelated errors we used the following three equation model which is a modification of the one used in Chapter 5.

$$Y\,B + Y_{-1}\Gamma_1 + X\,\Gamma_2 = U$$

$$U = U_{-1}R + E$$

$$E(\,E_t\,) = 0$$

$$E(\,E_t{}'E_t\,) = \Theta$$

where Y is a T.3 matrix of observation on the endogenous variables, Y_{-1} is T.3 matrix of observation on the lagged endogenous variables, X is a T.7 matrix of observation on the exogenous variables whose first element at every observation is unity, U is the T.3 observation matrix of the disturbances, R is a 3.3 diagonal matrix of the autocorrelation coefficients and E is T.3 matrix of random errors. B, Γ_1 and Γ_2 are the structural coefficients. To make our results comparable with the previous experiments, we used the same structure of the coefficients of the endogenous and the exogenous variables specified in Chapter 5. The structure of the coefficients of the lagged endogenous variables, Γ_1 is:[61]

----

[61]Given the structure of the coefficients, we can show that the dynamic system (1) is stable, i.e., the characteristic roots of $\Gamma_1 B^{-1}$ lie within a unit circle. For detailed discussion of the stability conditions of such dynamic systems see Murkata(1977).

$$\Gamma_1 = \begin{pmatrix} 0.25 & 0 & 0 \\ 0 & 0.29 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

We used both trended and nontrended stochastic exogenous variables. First, we examine the relative performance of the limited information estimators using non-trended stochastic exogenous variables. These variables follow the same process defined in Chapter V. Then, we investigate the effect of the trended data (to be specified later) on the performance of the estimators. The structure of the variance-covariance matrix of the errors, Θ, will be the same as in Chapter V.

As was the case in the previous experiments, we first undertook a pilot study which covered all the possible variations of the different estimators. Then we selected the estimators which performed relatively better than the others for the rest of the experiments. The pilot experiment undertaken in this part will investigate the performance of 9 estimators shown in Table 6-1, when the sample size is 30 and the autocorrelation coefficient of the first structural equation is equal to 0.6 while the second and third equations possess autocorrelation of 0.9 and 0.2 respectively. To increase the efficiency of the experiment without increasing the computational cost we employed the antithetic variate method. Each complete experiment consisted of 200 replications, 100 direct and 100 antitethic. For the experiments which use the effective sample size of 30, we generated 50 observations on X_t and

TABLE 6-1: Structure of the Estimators Used in the Dynamic Autoregressive Model

Augmented Reduced Form		Ordinary Reduced Form
T-1	T-2	T
Fair	Theil(G2SLSM)	Theil(G2SLSM)
Dhrymes (C2SLSA)		Dhrymes(2-Step)
Hatanaka(A)		Dhrymes(2-Step) (Iterated)
Hatanaka(B)		
Hatanaka(C)		

* Note that T is the number of observations.

U_t.[62] Using X_t and U_t, the Y_t series were constructed, using zero for the initial value of Y_{-1}. Then the first 20 observations were rejected to make the sample independent of the starting values. For those experiments which used the effective sample size of 60, we generated 100 observations on X_t and U_t and thus on Y_t, using the initial value of zero for Y_{-1}. Then the first 40 observations were rejected.

All estimators considered in this part (except Fair) have the first step, which is an instrumental variable estimation of the structural coefficients under consideration, in common. Therefore, for the first step IV estimation we use the fitted values of the endogenous variables estimated from the augmented reduced form as instrumental variables for the current and the lagged endogenous variables. For the Theil estimator which employs the ordinary reduced form, we chose $\hat{Y}_{-1} = X_{-1}\hat{\Pi}$ as the instrument for Y_{-1}.

For a finite sample size, instrumental variable estimators might lack moments of any order. This, as Hatanaka pointed out[63], is due to the fact that the matrix[64] $(W'Z)$ might not be positive definite and therefore the probability of the determinant of $(W'Z)$ being equal to zero can be non-zero which causes the integral that defines

---

[62] Note that effective sample size "T" means having "T+1" observations since the existence of lagged endogenous variables automatically reduces the number of observations by one.
[63] Hatanaka (1974), pp.205-206.
[64] Note that the standard IV estimator is in the form of: $B=(W'Z)^{-1}W'y$.

the moments of the IV estimator to become infinite. Since all estimators in this part are based on the first stage IV estimation, there is, thus, a possibility that they also lack moments of any order.[65] Therefore, to deal with this problem, we shall accompany our standard summary statistics with statistics on the median as a measure of centre, and the interdecile range, i.e., the distance between the quantiles of order 0.1 and 0.9, as a measure of dispersion.[66]

D. Monte Carlo Results Using Stochastic Non-trended Exogenous Variables

a. Results of the Pilot Experiment

The pilot experiment was carried out using the antithetic method to compare the performance of the estimators shown in Table 6-1. We encountered major problems in using three of the estimators, namely Theil which uses the ordinary reduced form, the Dhrymes two-step estimator and its iterative version. In many instances they failed to converge after 20 iterations. Moreover, they often produced unacceptable results(i.e., r>1) when they converged. Also,

---

[65] Note that this problem did not exist for the models without lagged endogenous variables. They were all based on a first step 2SLS estimator whose first two moments are known to exist for the system under consideration in the first part of this thesis. Moreover, we calculated the measures of centre and location that are not based on the existence of the moments for the non-lagged endogenous variable model and we did not find any change in the ranking of the estimators.

[66] These are the summary statistics used by Hatanaka (1974), p.206.

in few of the iterations for the Dhrymes two-step estimator we obtained singular matrices. Therefore, we omitted those estimators and carried out the experiment with the remaining six estimators.

b. General Results of the Monte Carlo Experiment

In addition to these six estimators, we considered 2SLS and the true Theil estimators. The 2SLS method estimates the structural coefficients with the restriction that the autocorrelation coefficient is equal to zero. This,in fact, is a misspecification. Due to the presence of lagged endogenous variables, this misspecification not only leads to inefficient results, but also gives inconsistent coefficient estimates. The true Theil estimator estimates the structural coefficients under the assumption of known autocorrelation coefficient and will be used as a base for comparison. Appendix V contains the detailed statistics concerning the relative performance of the above estimators. In what follows, however, we shall compare them according to their aggregate indices.

1. Comparisons using measures of centre

Table 6-2 presents the results of the antithetic estimates of the aggregate biases of different estimators for different sample sizes. For low and medium degrees of autocorrelation, 2SLS has the lowest bias. Its bias

TABLE 6-2: Aggregate Bias Calculated on the Base of the Mean*

Autocorrelation Coefficient	T	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
0.2	30	0.69	0.80	0.63	0.26	0.55	0.66	0.28
	60	0.43	0.44	0.42	3.06	0.44	0.43	0.13
0.6	30	0.54	1.05	0.54	2.10	0.59	0.54	0.24
	60	0.33	0.64	0.32	0.31	0.55	0.31	0.08
0.9	30	0.54	1.52	0.73	0.19	1.09	0.56	0.74
	60	0.29	1.53	0.49	0.47	1.26	0.27	1.02

*Note this index does not include the intercept.

increased sharply as the autocorrelation coefficient reached 0.9. Theil, Hatanaka(A) and Dhrymes estimators have more or less the same aggregate bias. As was the case in the non-lagged system, the Fair estimator has the highest aggregate bias. As we discussed before, the index of the aggregate bias does not include the constant term. When the autocorrelation coefficient reached 0.9, all estimators seriously underestimated the constant term. This underestimation can cause serious problems for prediction purposes. The aggregate bias of Theil, Hatanaka(A) and Dhrymes estimators decreased considerably when the sample size was increased to 60, revealing their consistency property. The aggregate bias of 2SLS increased considerably when the autocorrelation coefficient was equal to 0.9 and the sample size was increased to 60. However, this behaviour which is in accordance with its inconsistency property only shows up when the autocorrelation coefficient is very high. For low and medium degrees of autocorrelation 2SLS had the lowest bias. Generally, the Theil, Hatanaka(A) and Dhrymes estimators performed best when the sample size increased and the autocorrelation coefficient was high. The above picture remains more or less the same when we use median rather than the mean as a measure of location (Tables 6-3). This correspondance of median and mean suggests that distribution of these estimators is symmetric.

2. Comparison using Normalized Determinant(MSE) Criterion

TABLE 6-3: Aggregate Bias Calculated on the Base of the Median*

Autocorrelation Coefficient	T	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
0.2	30	0.70	0.83	0.64	0.22	0.57	0.66	0.27
	60	0.43	0.50	0.42	0.06	0.51	0.44	0.11
0.6	30	0.55	0.96	0.53	0.39	0.61	0.54	0.23
	60	0.33	0.57	0.32	0.40	0.45	0.31	0.08
0.9	30	0.53	1.48	0.72	0.25	1.06	0.54	0.71
	60	0.28	1.55	0.46	0.56	1.27	0.26	1.14

*Note this index does not include the intercept.

A more revealing picture of the performance of different estimators can be observed by examining Table 6-4 in which we have normalized the det(MSE) of the estimators by dividing it by the corresponding det(MSE) of the true Theil estimator. Using the normalized det(MSE) as an aggregate measure of relative efficiency of different estimators, we observed that for sample size of 30 the Hatanaka(A) estimator performed best for low degrees of autocorrelation. Its performance deteriorated marginally when the autocorrelation coefficient was increased to 0.6 and became inferior to the Dhrymes and Theil estimators for high degrees of autocorrelation. Dhrymes's estimator ranks best for medium and high degrees of autocorrelation. Generally, three of the estimators, Dhrymes, Theil and Hatanaka(A) performed considerably better than all other estimators for all degrees of autocorrelation when the sample size was equal to 30. The performance of 2SLS worsened as the autocorrelation coefficient increased. The Fair estimator performed poorly relative to others. This picture remained more or less the same when the sample size was increased to 60. In this case, the relative position of the Theil and Hatanaka(A) estimator changed slightly; the Theil estimator performed best for low degree of autocorrelation. Dhrymes' estimator performed best for medium and high degrees of autocorrelation. The Hatanaka(A) estimator performed poorly relative to Dhrymes and Theil when the autocorrelation coefficient reached 0.9.

TABLE 6-4: Normalized Determinant(MSE)*

Autocorrelation Coefficient	T	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
0.2	30	2.15	5707	1.46	5.4x10	236.10	2.63	4.88
	60	4.25	82x10	5.53	1.3x10	3333.33	4.36	10.77
0.6	30	3.74	970833	3.08	43.0x10	240.42	2.49	63.33
	60	2.87	19x10	3.58	1.1x10	33720.9	2.70	12790.7
0.9	30	1.91	250411	16.14	17.0x10	3336.08	1.57	1632.6
	60	1.25	36x10	653.50	1.1x10	2x10	0.83	25x10

*Note this index does not include the intercept.

3. Comparison Using Trace(MSE) as a Criterion

A different picture appears when we compare estimators on the basis of the normalized trace(MSE) (Table 6-5). For sample size 30, 2SLS dominated all other estimators for low and medium degrees of autocorrelation. The Hatanaka(C) estimator also performed relatively well for low degrees of autocorrelation. The relative performance of 2SLS and Hatanaka(C) deteriorated as the autocorrelation coefficient reached 0.9. For high degree of autocorrelation Theil, Dhrymes and Hatanaka(A) estimators performed best. The relative ranking of the estimators changed slightly when the sample size increased to 60. 2SLS ranked best only for low degree of autocorrelation. Dhrymes, Theil and Hatanaka(A) estimators performed best for medium and high degrees of autocorrelation.

As we mentioned before, comparison of estimators on the basis of the normalized det(MSE) and trace(MSE) implicitly assumes the existence of the small sample moments. However, the estimators that are based on the first step instrumental variable estimation might not possess small sample moments. All the estimators, except 2SLS, considered in this part are on the basis of first step instrumental variable estimation. This might explain the relative superiority of 2SLS for low and medium degrees of autocorrelation when the sample size was equal to 30 and comparison was made on the basis of normalized trace(MSE).

TABLE 6-5: Normalized Trace(MSE)*

Autocorrelation Coefficient	T	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
0.2	30	1.10	2.25	0.94	73.82	0.81	1.02	0.31
	60	1.09	1.74	1.04	1153.11	1.80	1.09	0.42
0.6	30	0.99	11.11	0.91	1003.49	1.48	0.97	0.54
	60	0.99	6.48	0.98	158.93	4.21	0.98	1.16
0.9	30	1.16	11.84	1.88	57.98	4.56	1.22	3.28
	60	0.94	30.14	2.84	1470.29	15.35	0.84	14.11

*Note this index does not include the intercept.

4. Comparison Using Interdecile Range as a Criterion

Table 6-6 shows the normalized trace of the interdecile range of different estimators for sample sizes of 30 and 60. Analogously to the normalized trace(MSE), we have normalized the sum of the interdecile range across coefficients for each estimator. To make the results comparable to the trace(MSE), we have omitted the intercept term from the calculation of this index. Using this index as a criterion of evaluation we can observe that for a low degree of autocorrelation Hatanaka(A) performed marginally best while for medium and high degrees of autocorrelation the Theil and Dhrymes estimators outperformed all others. The performance of the Hatanaka(A) estimator deteriorated marginally as the autocorrelation coefficient increased. The 2SLS estimator did not perform well even for low degrees of autocorrelation. Its relative performance deteriorated as the autocorrelation coefficient increased. This behavior is in accordance with its inconsistency and inefficiency properties. Hatanaka(C) performed relatively well and ranked fourth for medium and high degrees of autocorrelation when the sample size was 30. The above picture did not change when the sample size was increased to 60.

It is noteworthy that the above ranking of the estimators is more or less the same as the one that emerged using the normalized det(MSE).

To summarize this section, the ranking of estimators depends on the criteria that one uses to measure their

TABLE 6-6: Normalized Trace(Interdecile Range)*

Autocorrelation Coefficient	T	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
0.2	30	1.16	3.18	1.15	13.71	1.66	1.18	1.63
	60	1.12	2.65	1.04	10.42	1.69	1.13	1.18
0.6	30	1.11	3.48	1.13	13.18	1.56	1.09	1.87
	60	1.00	3.70	1.07	14.99	1.38	1.04	1.85
0.9	30	1.09	4.78	1.49	9.32	2.08	1.13	2.95
	60	1.00	5.72	1.89	14.81	1.04	1.01	4.46

*Note this index does not include the intercept.

relative efficiency. The ranking made using the interdecile range is theoretically more sound since employment of the measures that are based on the existence of small sample moments can be misleading in the present case. Therefore, we can conclude that the Theil, Dhrymes and Hatanaka(A) estimators performed best among all the potential estimators for the dynamic simultaneous equation models with autocorrelated errors. There exist a significant gain of efficiency in using these methods which take account of autocorrelation and the lagged endogenous variables, even for very low degrees of autocorrelation. 2SLS performed uniformly poorly relative to these estimators even when the autocorrelation coefficient was equal to 0.2.

An important observation is that the Fair estimator which is the most commonly known method for estimation of dynamic autoregressive models and is incorporated in the TSP package, was inferior to the Theil, Hatanaka(A) and Dhrymes estimators. Therefore, the above results suggest a potential danger in using the Fair estimator for estimation of the dynamic autoregressive models. Note also that Hatanaka(B) which is based on a Brundy and Jorgenson method generally does poorly which is in agreement with the results in the static model.

E. Monte Carlo Results Using Stochastic Trended Exogenous Variables

To examine the effect of trended data on the performance of dynamic autoregressive estimators we used the following exogenous variables

$$X_1 = e^{0.15t} + w_1 \qquad w_1 \sim N(0,\ 0.0049)$$

$$X_2 = e^{0.12t} + w_2 \qquad w_2 \sim N(0,\ 0.0025)$$

$$X_3 = e^{0.10t} + w_3 \qquad w_3 \sim N(0,\ 0.0009)$$

X_4, X_5 and X_6 are defined to be stochastic non-trended as in Chapter 5.

To minimize the cost, we only conducted experiments using a sample size of 30 and autocorrelation coefficient of 0.6. We omitted the Hatanaka(B) and Hatanaka(C) estimators, since they performed poorly in the previous case. The reason for introducing the Fair estimator is for the purpose of making our results comparable to the ones obtained in the first part of this study.

1.Comparison Using Measures of Centre

Table 6-7 shows the results of the antithetic estimates of the biases of different estimators for the trended data. As can be seen the 2SLS has the lowest aggregate bias followed by the Dhrymes and Theil estimators. In estimation of the autocorrelation coefficient, the Theil estimator has the lowest bias followed by the Dhrymes estimator. The Fair estimator has the highest aggregate bias. In estimation of the constant term which is not includedin the index of

TABLE 6-7: Bias Calculated on the Basis of the Mean Using Trended Data
(r=0.6, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Dhrymes	2SLS
C	-2.39	-12.66	2.57	2.74	-2.94
B_1	0.25	0.34	0.26	0.25	0.22
B_2	-0.13	-0.42	-0.18	-0.13	-0.24
B_3	-0.23	-0.25	-0.23	-0.23	-0.18
C_1	-0.10	-0.20	-0.10	-0.09	-0.11
C_2	-0.46	-0.55	-0.45	-0.45	-0.36
r	-0.12	-0.42	-0.26	-0.15	N.A
Aggregate* Bias	0.61	0.93	0.66	0.60	0.53

*This index does not include the intercept term. Note that the structural equation under consideration in this chapter is

$$y_1 = Y_1 B_1 + Y_2 B_2 + y_{1,-1} B_3 + X_1 C_1 + X_4 C_2 + u_1$$

aggregate bias, the Fair estimator performed worst. It seriously underestimates the intercept term. The rest of the estimators had more or less the same absolute bias in the estimation of the intercept term.

The above picture does not change when we consider the median rather than the mean as a measure of location (Table 6-8). The biases of the estimators stayed more or less unchanged. This correspondence of the mean and median reflects the symmetric property of these estimators.

The above pattern of the aggregate bias of the estimators using trended data is more or less the same as the one that emerged using non-trended data (Tables 6-2, 6-3). In other words, the existence of trend in the exogenous variables did not significantly change the relative performance of the estimators.

2. Comparison Using Normalized Determinant(MSE) Criterion

Table 6-9 presents the normalized MSE and det(MSE) of different estimators for trended data. Ranking the estimators on the basis of their normalized det(MSE), we can observe that Hatanaka's Residual Adjusted estimator performed best. The Dhrymes and Theil estimators ranked second and third. The 2SLS estimator was outperformed by the above three estimators. The Fair estimator performed poorly. Its performance in this case is in contrast to its performance in the static model. In other words, Fair's performance in the model with lagged endogenous variables,

TABLE 6-8: Bias Calculated on the Basis of the Median Using Trended Data
($r=0.6$, $T=30$)

Coefficient	Theil	Fair	Hatanaka(A)	Dhrymes	2SLS
C	1.41	-12.43	2.21	3.13	-3.05
B_1	0.25	0.32	0.25	0.25	0.21
B_2	-0.13	-0.38	-0.17	-0.13	-0.22
B_3	-0.23	-0.24	-0.23	-0.23	-0.17
C_1	-0.10	-0.19	-0.10	-0.09	-0.10
C_2	-0.45	-0.51	-0.44	-0.44	-0.36
r	-0.09	-0.38	-0.25	-0.13	N.A
Aggregate* Bias	0.59	0.86	0.64	0.59	0.51

*This index does not include the intercept term.

TABLE 6-9: Normalized Root Mean Squares Error Using Trended Data
($r=0.6$, $T=30$)

Coefficient	Theil	Fair	Hatanaka(A)	Dhrymes	2SLS
C	1.99	1.89	1.04	1.47	1.42
B_1	1.02	1.57	1.06	1.03	1.07
B_2	1.36	4.01	1.56	1.34	2.55
B_3	1.00	1.16	1.00	1.00	0.84
C_1	1.51	2.65	1.28	1.33	1.78
C_2	1.02	1.34	0.99	1.01	0.91
Normalized* Det(MSE)	26.54	1244.59	4.84	11.28	46.00
Normalized* Trace(MSE)	1.11	2.77	1.11	1.09	1.24

*This index does not include the intercept term.

for which it was specifically designed, was not as strong as in the static case when the variables were trended.

The ranking of the Hatanaka and Theil estimators has changed slightly due to the existence of trend in the exogenous variables. However, the general picture is more or less the same as the one obtained using stochastic non-trended exogenous variables.

3. Comparison Using Trace(MSE) as a Criterion

The last row of Table 6-9 gives the normalized trace(MSE) for different estimators. Using this index as a criteria of evaluation, we can see that the relative ranking of Hatanaka and Dhrymes estimators changed compared to the one obtained using GMSE. However, the general pattern of the relative performance stayed unchanged. The Dhrymes , Hatanaka and Theil estimators outperformed all others.

In the case of stochastic non-trended exogenous variable (Table 6-5) we observed that the ranking of the estimators changed when we used trace(MSE) instead of the normalized det(MSE). In that case the 2SLS performed best on the basis of the normalized trace(MSE). However, in the present case of stochastic trended data, the relative ranking of the estimators did not change. In other words, the performance of the 2SLS estimator was not affected by the measure of goodness used. This might suggest that the existence of trend in the exogenous variables has a negative effect on the performance of the 2SLS, judged on the basis

of the normalized trace(MSE).

4. Comparison Using Interdecile Range as a Criteria

Table 6-10 shows the normalized interdecile range of different estimators using trended data. The last row of this Table shows the normalized trace of the interdecile ranges. As we mentioned before, this index does not include the intercept term.

Comparing estimators on the basis of their trace interdecile range gives the same ranking as the one obtained using trace(MSE). This relative ranking of the estimators using trended exogenous variables is the same as the one emerged using stochastic non-tranded variables.

To summarize, the ranking of the estimators did not significantly change due to the existence of trend in the exogenous variables. This observation is in line with the results obtained in the first part using models without lagged endogenous variables.

F. Performance of Estimators in Hypothesis Testing

Detailed statistics concerning the Type I error and the power of the tests are given in Appendix VII. In this section we present the summary indices used in Chapter V. Table 6-11 and 6-12 show the average Type I error and the the average power of the test of significance of different estimators for different sample sizes and autocorrelation coefficients when the exogenous variables are stochastic

TABLE 6-10: Normalized Interdecile Range Using Trended Data
(r=0.6, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Dhrymes	2SLS
C	1.65	1.96	1.11	1.13	1.45
B_1	1.25	3.79	1.22	1.12	3.10
B_2	1.56	4.27	1.53	1.42	2.71
B_3	1.06	2.98	1.17	1.17	2.55
C_1	1.26	2.37	1.06	1.01	1.72
C_2	1.20	3.53	1.23	1.21	2.94
Normalized* Trace	1.30	3.42	1.26	1.20	2.56

*This index does not include the intercept term.

TABLE 6-11: Average Type I Error at 5% Significant Level

Autocorrelation Coefficient	T	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
0.2	30	69 (6-96)	34 (3-51)	63 (4-88)	30 (1-55)	68 (5-95)	23 (4-34)
	60	58 (10-74)	31 (1-48)	56 (11-74)	38 (2-68)	59 (10-76)	18 (10-25)
0.6	30	56 (10-80)	35 (7-51)	49 (7-75)	41 (2-76)	55 (10-80)	22 (8-29)
	60	42 (9-60)	32 (9-41)	39 (13-58)	56 (2-89)	42 (10-63)	12 (10-28)
0.9	30	36 (8-53)	30 (5-40)	41 (5-63)	49 (1-80)	39 (9-60)	35 (9-56)
	60	25 (15-34)	40 (28-54)	37 (29-64)	77 (23-99)	26 (16-34)	58 (40-67)

TABLE 6-12: Average Power of Test of Significance at 5% Significant Level

Autocorrelation Coefficient	T	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
0.2	30	46 (6-100)	38 (5-89)	46 (3-100)	28 (4-94)	46 (6-100)	47 (4-95)
	60	61 (6-100)	37 (2-92)	60 (5-100)	41 (17-100)	61 (7-100)	66 (7-100)
0.6	30	48 (5-100)	37 (1-92)	46 (5-100)	34 (5-99)	48 (5-100)	46 (9-91)
	60	70 (20-100)	37 (8-90)	70 (17-100)	42 (18-100)	72 (19-100)	66 (15-93)
0.9	30	48 (3-100)	37 (10-85)	44 (9-100)	39 (5-98)	48 (7-100)	44 (16-94)
	60	67 (8-100)	36 (10-93)	54 (14-100)	58 (4-100)	71 (8-100)	58 (28-97)

non-trended.[67] The figures in paranthesis show the range of the corresponding statistics over the structural coefficients. These tables show that for the sample size of 30 the observed level of Type I error for all estimators is much greater than the nominal level of significance for all degrees of autocorrelation. This problem was specifically clear in the case of Theil, Hatanaka(A) and Dhrymes' estimators for low and moderate autocorrelation coefficients. When the sample size was increased to 60 the observed Type I error of most of the estimators decreased. However, it was still much larger than the nominal level of significance. Table 6-12 shows the power of the test of significance for different estimators. All estimators performed poorly. These results resemble the ones obtained by Park and Mitchell(1981) in the single equation context. They attributed the high percentage of Type I errors to the underestimation of the standard errors. The above picture remained more or less unchanged when we used stochastic trended exogenous variables (Tables 6-13, 6-14).

The above results lead us to the same conclusion reached by Park and Mitchell (1981, p.199) that " distrust the conventional t-statistics....Because estimated coefficients seem much more significant than they really are, apply a more stringent confidence level for hypothesis testing."

----

[67] It must be noted that the Hatanaka(B) estimator was excluded from this section since, in number of instances, it gave rise to negative variances that were not acceptable.

TABLE 6-13: Percentage Type I Error at 5% Significant Level Using Trended Data
($r=0.6$, $T=30$)

Coefficient	Theil	Fair	Hatanaka(A)	Dhrymes	2SLS
C	18	21	27	21	25
B_1	94	65	91	95	55
B_2	19	36	26	19	33
B_3	97	73	97	98	57
C_1	34	41	36	32	32
C_2	85	55	74	87	37
Average	66	54	65	66	43

TABLE 6-14: Percentage Power of the Test of Significance at 5% Level Using Trended Data
(r=0.6, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Dhrymes	2SLS
C	46	42	62	54	45
B_1	100	97	100	100	97
B_2	26	29	12	24	18
B_3	6	10	11	8	17
C_1	94	77	99	98	93
C_2	4	4	4	5	5
Average	46	43	45	47	46

These results suggest that the test of hypothesis can be a serious problem for the dynamic autoregressive models. In Chapter V we saw that some of the estimators performed well in hypothesis testing when the model did not include lagged endogenous variables. However, the results of this section shows that none of the estimators can be reliable in test of hypothesis in the dynamic autoregressive models.

VII. SUMMARY AND CONCLUSION

Time series data often generates disturbances which are time dependent. The simplest form of this time dependence is a first order autocorrelation. Therefore, it is of great interest for the econometricians working with time series data to inquire into the properties of methods appropriate for the estimation of models characterized by autocorrelation.

Numerous techniques have been proposed for estimation of first order autocorrelated models in simultaneous equation systems. These methods can be distinguished by whether they use T or T-1 observations, the reduced form they employ and the way they estimate the autocorrelation coefficient. Most of the proposed methods are asymptotically efficient but nothing is known about their small sample properties. This thesis was an atempt to fill this gap using Monte Carlo methods.

We chose the three equation model used by Cragg(1967,1968) for our experiments. The choice of a three equation model was on the basis of the Mosbaek's(1970) findings suggesting that three equation models reveal the essential properties of larger models whereas two equation models do not. To increase the efficiency of our experiments we chose the antithetic variate method which is more efficient than the conventional direct simulation method. Our pilot experiment showed that on average we achieved a gain in efficiency of about 20 percent. Each complete

experiment consisted of 200 replications, 100 direct and 100 antithetic. We concentrated on the limited information methods for reasons of cost and relevance.

Our investigation consisted of two major parts. The first part examined the small sample properties of the estimators designed for estimation of the autocorrelated models without lagged endogenous variables. The second part studied the properties of estimators appropriate for estimation of dynamic autoregressive models.

Findings from the Static Model

In the first part, we considered almost all the limited information methods proposed in the literature. We also suggested new estimators and different ways of interpreting or deriving the existing methods. To minimize the probability of exclusion of any potentially good estimator, we first conducted a pilot experiment and then chose the estimators that performed relatively well for the subsequent experiments. We conducted experiments with both trended and non-trended data.

Recently Maeshiro(1976,1979,1980) and Park and Mitchell(1980) compared the relative efficiency of single equation estimators that use all the observations with those that omit the first observation. They found that the estimators which omit the first observation are inferior to the ones which use all observations, especially when the data was trended. Following them, we compared the relative

efficiency of the limited information estimators that use T or T-1 observations. We found that employment of an extra observation had some positive effect on the small sample performance of estimators, especially when the exogenous variables were trended. However, the difference was not as great as was suggested by Maeshiro(1976,1879) and Park and Mitchell(1980). To investigate the apparent discrepancy of the results, we undertook a Monte Carlo experiment replicating and extending their investigation. We found that the source of the discrepancy was the restricted nature of the model employed by them.

Comparison of the methods employing different reduced form showed that the augmented reduced form produced more efficient estimates. We also found that the Prais-Winsten method for estimation of the autocorrelation coefficient always leads to the estimate that has lower bias and RMSE. It should be noted that the Theil estimator had the lowest bias in estimation of the autocorrelation coefficient, even when it was using the CORC formula, i.e., in the dynamic autoregressive case.

A more important finding of this study concerns with relative performance of different methods. In practice, the Fair estimator is the most commonly known method for estimation of the models characterized by autocorrelation. Our results showed that (except for some cases when the exogenous variables were trended) this estimator was highly inefficient relative to alternatives we considered. In some

instances, it was even inferior to the 2SLS which ignores autocorrelation. We also found that the small sample bias of 2SLS estimator was positively related to the degree of autocorrelation.

The ranking of different estimators turned out to be slightly dependent on the criteria of evaluation used. For a low autocorrelation coefficient of 0.2, 2SLS outperformed all other estimators on the basis of generalized MSE, while it was not best if we had used trace(MSE) as a measure of goodness. Fair and modified Brundy and Jorgenson along with the Amemiya and generalized limited information maximum likelihood estimators performed poorly. Generally, the Theil estimator that has been completely ignored in empirical research turned out to be the best method for autocorrelated models without lagged endogenous variables.

At this point we should point out that in large models we may not be able to use Theil's estimator due to the degrees of freedom problem. In fact, the Fair estimator is specifically designed to solve this problem by using fewer instruments than estimators such as Theil. However, its weak performance cast some doubt on its usefulness. As an alternative one may use some variation of Theil's estimator which uses, for instance, a subset of the principal components of the exogenous variables.

Following Park and Mitchell (1981), we examined the performance of different estimators in the test of hypothesis. Following them we focused on the number of times

each method leads to the occurrence of a Type I error at 0.05 level of significance. We found that perhaps the main problem with employment of 2SLS and Fair estimators, in the case of static models and the sample size of 30, is that they lead to a much greater Type I error than the nominal level of significance suggests when the autocorrelation coefficient is moderate or high. This fact suggests that the tail of their test statistic distributions are much thicker than those given by the t-distribution. The Theil estimator outperformed all other estimators for moderate and high autocorrelation coefficients. As the sample size increased to 60, all estimators experienced a significant reduction in their actual Type I errors. In fact, their observed Type I errors became very close to the nominal level of significance.

Generally, our results lead us to the conclusion that for very low autocorrelation coefficients, 2SLS is as good as any other estimator in static models with autocorrelated disturbances. However, as the autocorrelation coefficient increases the performance of 2SLS deteriorates. The Theil estimator and its modified version perform best for moderate and high autocorrelation coefficients when the model does not include lagged endogenous variables.

Findings from the Dynamic Model

The second part of this thesis dealt with dynamic autoregressive models. Very little is known about the small

sample performance of the methods appropriate for estimation of these kinds of models. We considered almost all the existing methods. We also extended the Theil estimator to take into account lagged endogenous variables. Apart from the conventional criteria of goodness, we also provided information on the median and interdecile range for all estimators.

As was the case in the static models, the ranking of the estimators was sensitive to the criteria of evaluation used. Focusing on the interdecile range for theoretical reasons, we found that Theil's, Dhrymes' and Hatanaka's Residual Adjusted estimators performed best. The performance of Hatanaka's Residual Adjusted estimator deteriorated slightly as the autocorrelation coefficient increased. Two stage least squares did not perform well even for low degrees of autocorrelation. Its performance, however, deteriorated as the autocorrelation coefficient increased. This behavior is consistent with its asymptotic property. Fair along with the other two Hatanaka estimators did not perform well. Our investigation showed that the Brundy and Jorgenson type estimators performed relatively poorly. In fact, we found that none of the estimators that calculated the fitted values of the endogenous variables, using consistent estimates of the reduced form coefficient performed well.

The effect of trended data on the performance of dynamic autoregressive estimators was examined using

stochastic trended exogenous variables. We found that in general the ranking of the estimators did not change and stayed the same as the one emerged using stochastic non-trended data.

We saw in the first part that the Fair estimator is asymptotically less efficient than the Theil estimator. One of the reason for its use is the argument that using fewer instruments at the initial stage might improve its small sample performance. However, our results showed the contrary. We also found that 2SLS which ignores autocorrelation generally performed poorly relative to other estimators. These results endorse Hendry's suggestion that the asymptotic properties can be used as guide to the small sample performance of different estimators. However, it should also be noted that ALIML which was shown to be asymptotically more efficient than the Theil's estimator when the autocorrelation coefficients across equations were not equal, did not perform well in small samples. This fact, in turn supports the Maasoumi and Phillips(1980)'s argument that the discrepancies between asymptotic and finite sample behavior of the estimators are parameter dependent and we cannot, without qualification, postulate the asymptotic behavior of the estimators from the small sample results.

The second part also examined the performance of different estimators in the test of hypothses. We found that for the dynamic autoregressive model, none of the estimators performed well. In fact all of them showed much greater Type

I error than the nominal level of significance. This result is in line with the Park and Mitchell's result in the single equation context and therefore leads us to the same type of conclusion that we can not trust the conventional t-statistics in the dynamic autoregressive models. Moreover, since the estimated coefficients seem to be much more significant that they really are, we should apply a much more stringent confidence level for hypothesis testing.

To summarize, our study supports the Rao and Griliches(1969) findings in the single equation context that "there is a significant gain in efficiency to be had from using two stage estimation procedures for moderate and high level of serial correlation in the residuals ($r>0.3$) and very little loss using such methods even when the true r is small". We also found that there is some gain in efficiency for those estimators that use all T rather than T-1 observations.

The results of this study can clearly be of value in guiding econometric practitioners in their choice of technique for estimation of models characterized by autocorrelation. However, our results are subject to the limitations inherent in all Monte Carlo studies. One of the ways that this study can be extended is the incorporation of Sargan's 2SLS estimator that is asymptotically equivalent to ALIML. The main reason for the omission of this estimator was that the results of the efficiency of ALIML relative to Theil's estimator became known to us when all the

experiments were done. Moreover, we showed that ALIML, GLIML and Theil's G2SLS estimators are equally efficient when the autocorrelation coefficients are equal across equations. Therefore, it would be interesting to compare these estimators in an experiment in which the autocorrelation coefficients are equal. It is conceivable, however, that the Theil estimator which outperformed other estimators when the autocorrelation coefficients were different across equations would continue to dominate when the autocorrelation coefficients are equal. This is so since the asymptotic superiority of ALIML was obtained when the autocorrelation coefficients were not equal across equations. Finally, a clear direction worthy of further research can be the extension of our investigation to the full information estimators designed for estimation of simultaneous equation models with autocorrelated errors.

BIBLIOGRAPHY

Amemiya, T. "Specification Analysis in the Estimation of Parameters of a Simultaneous Equations Models With Autocorrelated Residuals." *Econometrica* 34, No.2 (1966): 283-306.

Anderson, T. W. and H. Rubin "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Annals of Mathematical Statistics* 20 (1949): 46-63.

Basmann, R.L. "On the Asymptotic Distribution of Generalized Linear Estimators." *Econometrica* 28 (1960): 97-107.

Basmann, R.L. "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Overidentified Cases." *Journal of The American Statistical Association* 56 (1961):619-636.

Beach, C.M. and James G. MacKinnon "A Maximum Likelihood Procedure For Regression With Autocorrelated Errors." *Econometrica* 46, No.1 (1978):51-58.

Brundy, J. M. and D.W. Jorgenson "Efficient Estimation of Simultaneus Equations By Instrumental Variables." *Review of Economics and Statistics* 53, No.3 (1971):207-224.

Brundy, J. M. and D. W. Jorgenson "The Relative Efficiency of Instrumental Variables Estimators of Systems of Simultaneous Equations." *Annals of Economic and Social Measurement* (1974): 679-700.

Brundy, J. M. and D. W. Jorgenson "Consistent and Efficient Estimation of Systems of Simultaneous Equations By Means of Instrumental Variables." in Paul Zarembka, *Frontiers in Econometrics*, 1974, 215-244.

Buse, A. "Autocorrelation and Two Stage Least Squares: Variations on a Theme by Theil." *Research Paper* No.83-10 (1983).

Chipman, J. S. "Efficiency of Least Squares Estimation of Linear Trend when Residuals are Autocorrelated." *Econometrica* 47 (1979): 115-128.

Cragg, J.G. "Small Sample Performances of Various Simultaneous Equation Estimators in Estimating the Reduced Form." *Metroeconomica* 19 (1967a): 77-93.

Cragg, J.G. "On the Relative Small Sample Properties of Structural Equation Estimators." *Econometrica* 35, No.1 (1967b): 89-111.

Cragg, J.G. "Some Effects of Incorrect Specification on the Small Sample Properties of Several Simultaneous Equation Estimators." *International Economic Review* 9, No.1 (1968): 63-86.

Deistler, M. "Z-Transformation of Linear Econometric Models With Autocorrelated Errors." *Metrika* 22 (1975): 13-25.

Deistler, M. "The Identifiability of Linear Econometric Models With Autocorrelated Errors." *International Economic Review* 17 (1976): 26-45.

Deistler, M. and J. Schrader "Linear Models With Autocorrelated Errors: Structural Identifiability in the Absence of Minimality Assumptions." *Econometrica*. 47 (1979): 495-504.

Dhrymes, P. J. "Alternative Asymptotic Tests of Significance and Related Aspects of 2SLS and 3SLS Estimated Parameters." *Review of Economic Studies* 36 (1969): 213-226.

Dhrymes, P.J. *Econometrics: Statistical Foundation and Applications*. New York: Harper and Row Publishers, Inc, 1970.

Dhrymes, P. J. and R. Berner and D. Cummins " A Comparison of Some Limited Information Estimators for Dynamic Simultaneous Equations Models With Autocorrelated Errors." *Econometrica* 42 (1974): 311-332.

Dhrymes, P.J. and H. Erlat "Asymptotic Properties of Full

Information Estimators in Dynamic Autoregressive Simultaneous Equation." *Journal of Econometrics* 2 (1974): 247-259.

Doran, H. E. and W. E. Griffiths "On the Relative Efficiency of Estimators Which Include the Initial Observations in the Estimation of Seemingly Unrelated Regressions With First Order Autoregressive Disturbances." *Department of Economic Statistics*, University of New England, Armidale NSW 235, Australia, 1982.

Fair, R. "The Estimation of Simultaneus Equation Models With Lagged Endogenous Variables and First Order Serially Correlated Errors." *Econometrica* 38, No.3 (1970): 507-516.

Fair, R. "Efficient Estimation of Simultaneus Equations With Autoregressive Errors by Instrumental Variables." *Review of Economics and Statistics* 54 (1972): 444-449.

Fair, R. *A Short Run Forcasting Model of the United States Economy*. Massachusetts: D.C. Health and Company, 1971.

Fisher, F. M. *The Identification Problem in Econometrics*. New York: McGraw-Hill Book Company, 1966.

Fisher,G. R. "The Algebra of Estimation in Linear Econometric Systems." *International Journal of Mathematical Education in Science and Technology* 3, (1972):385-403.

Goldberger, A.S. "An Instrumental Variable Interpretation of K-class Estimation." *Indian Economic Journal* 13, No.3 (1965): 424-431.

Goldberger, A.S. *Econometric Theory*. New York: John Wiley and Sons, 1964.

Goldfeld, S.M. and R.E. Quandt *Nonlinear Methods in Econometrics*. Amsterdam: North-Holland Publishing Co., 1972.

Grenander, U. and M. R. Rosenblatt *Statistical Analysis of Stationary Time Series*. New York: John Wiley and Sons,

1957.

Haitousky, Y. and S. Jacobs "Regen-Computer Program to Generate Multivariate Observations for Linear Regression Equations." *Annels of Economic and Social Measurement* (1972): 43-57.

Hammersley, J. M. and D. C. Handscomb *Monte Carlo Methods*. London: Methuen and Co., Ltd., 1964.

Hammersley,J. M. and K. W. Morot "A New Monte Carlo Technique: Antithetic Variates." *Proceedings of the Cambridge Philosophical Society* 52 (1956): 449-475.

Hammersley, J. M. and J. G. Mauldon "General Principles of Antithetic Variates." *Proceedings of the Cambridge Philosophical Society* 52 (1956): 476-481.

Hannon, E. J. *Time Series Analysis*. London: Methuen and Co. Ltd, 1960.

Harvey, A. C. *The Econometric Analysis of Time Series*. Oxford: Philip Allan Publishers Limited, 1981.

Hatanaka, M. "An Efficient Estimator for the Dynamic Adjustment Model With Autoregressive Errors." *Journal of Econometrics* 2 (1974): 199-220.

Hatanaka, M. "Several Efficient Two-Step Estimators For the Dynamic Simultaneous Equations Model With Autoregressive Disturbances." *Journal of Econometrics* 4 (1976): 189-204.

Hendry, D.F. "Behaviour of the Inconsistent Instrumental Variable Estimators in Dynamic Systems With Autocorrelated Errors." *Journal of Econometrics* 9 (1979): 295-314.

Hendry, D.F. "The Structure of Simultaneous Equations Estimators." *Journal of Econometrics* 4 (1976): 51-88.

Hendry, D.F. "On Asymptotic Theory and Finite Sample Experiments." *Economica* 160 (1973):210-217.

Hendry, D.F. and P.K. Trivedi "Maximum Likelihood Estimation of Difference Equations With Moving Average Errors: A Simulation Study." *Review of Economic Studies* 39 (1972): 117-145.

Hendry, D.F. and R.W. Harrison "A Monte Carlo Methodology and the Small Sample Behaviour of Ordinary and Two-Stage Least Squares." *Journal of Econometrics* 2 (1974): 151-174.

Hendry, D.F. and F. Sbra "The Properties of Autoregressive Instrumental Variables Estimators in Dynamic Systems." *Econometrica* 45, No.4 (1977): 969-990.

Hildreth, C. and W. T. Dent "An Adjusted Maximum Likelihood Estimator of Autocorrelation in Disturbances." in Willy Sellekerts, *Econometrics and Economic Theory, Essays in Honour of Jan Tinbergen*. Toronto: Macmillan of Canada Limited, 1974.

Hsiao, C. "Identification", *Working Paper Series*, Institute for Policy Analysis, University of Toronto, No. 8103, 1981.

Hurd, M. D. "Small Sample Estimation of a Structural Equation With Autocorrelated Errors." *Journal of the American Statistical Association* 67 (1972): 567-573.

Johnston, J. *Econometric Methods*. Second Edition, New York: McGraw-Hill Book Company, 1972.

Judge, G. G., W. E. Griffiths, R. C. Hill and T. C. Lee *The Theory and Practice of Econometrics*. New York: John Wiley and Sons, 1980.

Kadane, J. B. "Comparison of K-Class Estimators When the Disturbances Are Small." *Econometrica* 39 (1971): 723-738.

Kelejian, H. H. and W. E. Oates *Introduction to Econometrics*. New York: Harper and Row Publishers, 1981.

Klein, L. R. *A Textbook of Econometrics*. London: Prentice-Hall International Inc, 1974.

Kmenta, J. *Elements of Econometrics*. New York: Macmillan Publishing Co., Inc., 1971.

Maasoumi E. and P. C. B. Phillips "On the Behaviour of Inconsistent Instrumental Variable Estimators." *Journal of Econometrics* 19 (1982):183-201.

Maddala, G. S. "Some Small Sample Evidence on Tests of Significance in Simultanous Equations Models." *Econometrica* 42 (1974): 841-851.

Madansky, A. "On the Efficiency of Three Stage Least Squares Estimation." *Econometrica* 32 (1964): 51

Maeshiro, A. "Autoregressive Transformations, Trended Independent Variables and Autocorrelated Disturbance Terms." *Review of Economics and Statistics* 58 (1976): 497-500.

Maeshiro, A. "On the Retention of the First Observation in Serial Correlation Adjustment of Regression Models." *International Economic Review* 20 (1979): 259-265.

Maeshiro, A. "New Evidence on the Small Sample Properties of Estimators of SUR Models With Autocorrelated Disturbances: Things Done Half-Way May Not Be Done Right." *Journal of Econometrics* 12 (1980): 177-188.

Mariano, R. S. "Analytical Small Sample Distribution Theory in Econometrics: The Simultaneous Equation Case." *International Economic Review* 23, No.3 (1982): 503-534.

Mosbaek, E.J. and H.O. Wold *Interdependent Systems; Structure and Estimation*. Amsterdam: North-Holland Publishing Company, 1970.

Mikhail, W.M. "Simulating the Small Sample Properties of Econometric Estimators." *Journal of the American Statistical Association* 67 (1972): 620-624.

Mikhail, W. M. "A Comparative Monte Carlo Study of the Properties of Econometric Estimators." *Journal of the American Statistical Association* 70 (1975): 94-104.

Murata, Y. *Mathematics for Stability and Optimization of Economic Systems*. New York: Academic Press, 1977.

Naylor, T. H. *Symposium on the Design of Computer Simulation Experiments*. Durham: Duke University Press, 1969.

Naylor, T. H. *Computer Simulation Experiments With Models of Economic Systems*. New York: John Wiley and Sons, 1971.

Nagar, A. L. "The Bias and Moment Matrix of the General K-Class Estimators of the Parameters in Simultaneous Equation." *Econometrica* 27 (1959): 575-595.

Nagar, A. L. "A Monte Carlo Study of Alternative Simultaneous Equation Estimators." *Econometrica* 28 (1960): 573-590.

Pagan, A. "Estimation and Control of Linear Econometric Models." *IHS-Journal* 6 (1982):247-268.

Park, R. E. and B. M. Mitchell "Estimating the Autocorrelated Error Model with Trended Data." *Journal of Econometrics* 13 (1980): 185-201.

Phillips, P. C. B. "Approximation to Some Finite Small Sample Distributions Associated With a First Order Stochastic Difference Equation." *Econometrica* 45 (1977): 463-486.

Phillips, P. C. B. "The Exact Distribution of Instrumental Variable Estimators in an Equation Containing n+1 Endogenous Variables." *Econometrica* 48 (1980): 861-878.

Phillips, P. C. B. "Finite Sample Theory and the Distributions of Alternative Estimators of the Marginal Propensity to Consume." *Review of Economic Studies* 67 (1980): 183-224.

Prais, G. and C. B. Winsten "Trend Estimators and Serial Correlation." *Discussion Paper* No.383, Cowles Commission, Chicago, Ill, 1954.

Pindyck, R. S. and D. L. Rubinfeld *Econometric Models and*

Economic Forecasts. New York: McGraw-Hill Book Company, 1981.

Rao, P. and Z. Griliches "Small Sample Properties of Several Two Stage Regression Methods in the Context of Auto-Correlated Errors." *Journal of the American Statistical Association* 64 (1969): 251-272.

Richardson, D. H. "The Exact Distribution of a Structural Coefficient Estimator." *Journal of the American Statistical Association* 63 (1968): 1214-1226.

Richardson, D. H. and R. J. Rohr "The Distribution of a Structural t-Statistic for the Case of Two Included Endogenous Variables." *Journal of the American Statistical Association* 66 (1971): 375-382.

Rubinstein, R. Y. *Simulation and the Monte Carlo Method*. New York: John Wiley and Sons, 1981.

Sargan, J. D. "The Estimation of Relationships With Auto-correlated Residuals By the Use of Instrumental Variables." *Journal of the Royal Statistical Society B-21* (1959): 91-105.

Sargan, J. D. "The Maximum Likelihood Estimation of Economic Relationships With Autoregressive Residuals." *Econometrica* 29 (1961): 414-426.

Sargan, J. D. and W. M. Mikhail "A General Approximation to the Distribution of Instrumental Variable Estimates." *Econometrica* 39 (1971): 131-169.

Sawa, T. "The Exact Finite Sampling Distribution of Ordinary Least Squares and Two Stage Least Squares Estimators." *Journal of The American Statistical Association* 64 (1969): 923-936.

Schink, W. A. and John S. Y. Chiu "A Simulation Study of Effects of Multicollinearity and Autocorrelation on Estimates of Parameters." *Journal of Financial and Quantitative Analysis* 1 (1966): 36-67.

Sellekaerts, W. *Econometrics and Economic Theory, Essays in*

Honour of Jan Tinbergen. Toronto: Macmillan of Canada Limited, 1974.

Smith, V. K. *Monte Carlo Methods: Their Role in Econometrics*. Massachusetts: D. C. Health and Company, 1973.

Sowey, E. R. "A Classified Bibliography of Monte Carlo Studies in Econometrics." *Journal of Econometrics* 1 (1973): 377-395.

Spitzer, J.J. "Small Sample Properties of Nonlinear Least Squares and Maximum Likelihood Estimation in the Context of Autocorrelated Errors." *Journal of the American Statistical Association* 74 (1979): 41-47.

Stewart and Wallis *Introductory Econometrics*. Basil Blackwell Publisher, 1981.

Summers, R. "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators." *Econometrica* 33 (1965): 1-41.

Taylor, W. E. "On the Efficiency of the Cochrane-Orcutt Estimators." *Journal of Econometrics* 17 (1981): 67-82.

Theil, H. *Economic Forecasts and Policy*. Amsterdam: North Holland Publishing Company, 1965.

Theil, H., *Principles of Econometrics*, New York: John Wiley and Sons, 1971.

Wang, G.H. and W.A.Fuller "Estimators for a Simultaneous Equation Model with Lagged Endogenous Variables and Autocorrelated Errors." *Paper Presented at the Third World Congress of the Econometric Society*, Toronto, Canada (1975): 20-26.

Wang, G.H. and W.A. Fuller "Estimators for a Simultaneous Equation Model with Lagged Endogenous Variables and Autocorrelated Errors." *Communication in StatisticsSimulation and Computation* 11, No.2 (1982):123-142.

Wickens, M. R. "The Consistency and Efficiency of Generalized Least Squares in Simultaneous Equation Systems With Autocorrelated Errors." *Econometrica* 37 (1969): 651-659.

Zarembka, P. *Frontiers in Econometrics*. New York: Academic Press, 1974.

Appendix I: DERIVATION OF THE LIML ESTIMATOR IN THE PRESENCE OF FIRST ORDER AUTOCORRELATION

In this appendix we follow closely the approach used by Koopmans and Hood(1953,pp.192-95). We also use their notation to make our derivations easily comparable to theirs.[68]

A. Sargan and Amemiya Version

Consider the following simultaneous equation system

$$B\ Y + \Gamma\ X = U \tag{1}$$

$$U = R\ U_{-1} + E$$

$$E_t = (\epsilon_{1t}, \ldots, \epsilon_{Gt})'$$

$$E(\epsilon_t) = 0$$

$$E(\epsilon_t \epsilon_t') = \Sigma$$

Following the practice in Amemiya(1961), Fair(1972), and Dhrymes(1972), we assume that R is a diagonal matrix of the form

$$R = \text{diag}(r_1,\ r_2, \ldots, r_G) \tag{2}$$

(1) can be transformed to the equation with a well behaved disturbances such as

$$B\ Y + \Gamma\ X - R\ B\ Y_{-1} - R\ \Gamma\ X_{-1} = E \tag{3}$$

or

$$A\ Z = E \tag{4}$$

where

$$A = (B,\ \Gamma,\ -RB,\ -R\Gamma)$$

----

[68] Note that this notation is only adopted in this appendix.

$$Z' = (Y', X', Y_{-1}', X_{-1}')$$

Under normality, the log-likelihood function of all the parameters of (4) is

$$L = K + T\log|B| - T/2 \log|\Sigma| - 1/2 \, tr(\Sigma^{-1}AMA') \qquad (5)$$

where M=Z Z' and Σ is the variance-covariance of E_t.

Since estimation of the first equation of (4) is the focus of our attention, we partition A as

$$A = \begin{pmatrix} A_1 \\ \\ A_2 \end{pmatrix} = \begin{pmatrix} B_1 & \Psi_1 \\ \\ B_2 & \Psi_2 \end{pmatrix} \qquad (6)$$

where A_1 is the first row of A, B_1 is the first row of B, Ψ_1 is the first row of (Γ, -RB, -RΓ) and A_2, B_2, and Ψ_2 are defined accordingly.

The corresponding partition of Σ is

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \qquad (7)$$

where Σ_{11} is 1.1, Σ_{12} is 1.(G-1), and Σ_{22} is (G-1)(G-1).

A_1 and Σ_{11} are parameters to be estimated and hence have to be retained, while A_2, $\Sigma_{12}=\Sigma_{21}'$, and Σ_{22} are parameters to be eliminated by partial maximization of (5). This can be achieved by first transforming the system (4) by a non-singular matrix F such that

$$F = \begin{pmatrix} 1_{11} & 0_1 \\ \\ F_{21} & F_{22} \end{pmatrix} \qquad (8)$$

where F_{22} is a (G-1).(G-1) non-singular matrix. This

transformation will leave the first equation unaffected. The transformed variance-covariance matrix is

$$\Sigma^* = F\,\Sigma\,F' = \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{pmatrix} \qquad (9)$$

where

$$\Sigma_{11}^* = \Sigma_{11} \qquad (10)$$

$$\Sigma_{21}'^* = \Sigma_{21}' = \Sigma_{11}F_{21}' + \Sigma_{12}F_{22}' \qquad (11)$$

$$\Sigma_{22}^* = F_{21}\Sigma_{11}F_{21}' + F_{22}\Sigma_{21}F_{21}' + F_{21}\Sigma_{21}F_{22}' + F_{22}\Sigma_{22}F_{22}' \qquad (12)$$

We want to choose F_{21} so that $\Sigma_{21}^* = 0$. This means that

$$F_{21}' = -\Sigma_{11}^{-1}\Sigma_{12}F_{22}' \qquad (13)$$

and substituting (13) into (12) yields

$$\Sigma_{22}^* = -F_{22}\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12} + F_{22}\Sigma_{22}F_{22}' \qquad (14)$$

$$= F_{22}[\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}]F_{22}'$$

We can use our freedom of choice of F_{22} so as to make $\Sigma_{22}^*=I_{22}$. This transformation not only makes the error terms of the last G-1 equations independent of the first equation but also makes them independent of each other and sets their variances equal to one.

In terms of the new parameters Σ^* and $A^*=\begin{pmatrix} A_1 \\ A_2^* \end{pmatrix}$, the matrix product under the trace of (5) becomes

$$\begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & I_{22} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2^* \end{pmatrix} M[A_1' \ A_2^{*\prime}] =$$

$$=\Sigma_{11}^{-1}A_1MA_1' + A_2^*M\,A_2^{*\prime} \qquad (15)$$

Using (15), the likelihood function (5) becomes

$$1/T\ L(A^*,\Sigma^*) = K_2+\log|B^*| -1/2\ \log|\Sigma_{11}|$$
$$-1/2\ tr(\Sigma_{11}^{-1}A_1MA_1') -1/2\ tr(A_2^*M\ A_2^{*\prime}) \qquad (16)$$

To obtain the concentrated likelihood function with respect to the elements of the first equation, we have to partially maximize (16) with respect to the elements of the second subset, A_2^*, and substitute the results back into (16).

The only terms containing the elements of the second subset in (16) are the second and the fifth terms. Therefore, partial maximization of (16) with respect to the elements of $A_2^*=(B_2^*,\Psi_2^*)$ will only affect those two terms. Now

$$\partial(\log|B^*|)/\partial b^*_{ij} = b^{*ji} = b_{ij}^{**}/|B^*| \qquad (17)$$
$$i=2,\ldots,G\ ;\ j=1,\ldots,G$$

Where b_{ij}^* is the element in the ith row and jth column of B^*, b^{*ji} is the element in the jth row and ith column of $(B^*)^{-1}$, and b_{ij}^{**} is the cofactor of b_{ij}^* and $|B^*|$ is the determinant of B^*.

Also we have

$$\partial\log|B^*| / \partial\Psi_{ik}^* = 0^{69} \qquad i=2,\ldots,G \qquad (18)$$
$$k=1,\ldots,G+2K$$

$$1/2\ [\partial tr(A_2^*M\ A_2^{*\prime})/\partial A_2^*]=1/2[2\ A_2^*M]=A_2^*M \qquad (19)$$

(17), (18) and (19) can be combined to give

$$\partial L/\partial A_2^* = \{[(B^{*\prime})^{-1}]_2\quad 0\} - A_2^{*\prime}M = 0 \qquad (20)$$

Note that A_2^* is a (G−1).(2G+2K) matrix of the coefficients of the second subset.

---

[69]Note that (18) holds only under the implicit assumption made by Sargan and Amemiya. We shall discuss and analyse this assumption later in this section.

From equation (20) we can calculate $A_2{*}M$ and using it to evaluate the fifth term in (16) such as

$$-1/2\ \mathrm{tr}(A_2{*}M\ A_2{*}') = -1/2\ \mathrm{tr}\{[(B{*}')^{-1}]_2\quad 0\}\begin{pmatrix} \beta_2^{*\prime} \\ \psi_2^{*\prime} \end{pmatrix}$$

$$= -1/2\ \mathrm{tr}\left(I_{T-1}\right) = \text{const.} \tag{21}$$

Note that the elements of the second subset of $(B*)^{-1}$, i.e., $(B*)_2^{-1}$ are equal to $b_{ij}{**}/|B*|$, where $b_{ij}{**}$ is the cofactor corresponding to the ith row and jth column of $|B*|$. Thus (21) holds since

$$[(B{*}')^{-1}]_2(B_2{*}') = (b{**}/|B*|)(B_2{*}') \tag{22}$$

where $b{**}$ is the matrix of the cofactors $b_{ij}{**}$. The ith row of the product (22) is equal to

$$(1/|B*|)[\Sigma\ b_{ij}{**}\ b_{ji}{*}] = (1/|B*|)(|B*|) = 1 \quad \text{if } i=j$$

$$= 0\ ,\ \text{if } i \neq j \tag{23}$$

Therfore, (22) is equal to

$$(1/|B*|)\begin{pmatrix} |B*| & 0 & 0 & \ldots & \\ 0 & |B*| & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & |B*| \end{pmatrix} = (1/|B*|)(|B*|)\begin{pmatrix} 1 & 0 & & \ldots \\ 0 & 1 & 0 & .. \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= I \tag{24}$$

Note that the sum of the product of each row of any matrix by its corresponding cofactors is equal to the determinant of that matrix, while the sum of the product of each row of any matrix by the cofactors corresponding to other rows is equal to zero.

To expand and evaluate the second term, $\log|B*|$, in (16) we partition the last term in (20) as

$$A_2{*}M = (B_2{*}\ \Psi_2{*})\begin{pmatrix} M_{yy} & M_{y\varphi} \\ M_{\phi y} & M_{\varphi\varphi} \end{pmatrix} \tag{25}$$

where

$$M = Z\ Z' = \begin{pmatrix} YY' & Y\Phi' \\ \Phi Y' & \Phi\Phi' \end{pmatrix} = \begin{pmatrix} M_{yy} & M_{y\phi} \\ M_{\Phi y} & M_{\Phi\varphi} \end{pmatrix}$$

where

$$\Phi' = (X'\ Y_{-1}'\ X_{-1}')$$

Thus we write (20) in expanded form as

$$[(B{*}')^{-1}]_2 = B_2{*}M_{yy} + \Psi_2{*}M_{\phi y} \tag{26}$$

$$0 = B_2{*}M_{y\varphi} + \Psi_2{*}M_{\varphi\varphi} \tag{27}$$

Calculating Ψ_2* from (27) as

$$\Psi_2{*} = -B_2{*}M_{y\varphi}M^{-1}_{\varphi\varphi}$$

and substituting it in (26) yields

$$[(B{*}')^{-1}]_2 = B_2{*}[M_{yy} - M_{y\varphi}M_{\phi\phi}{}^{-1}M_{\phi y}] = B_2{*}W \tag{28}$$

where

$$W = Y(I-\Phi'(\Phi\Phi')^{-1}\Phi)Y'$$

Now write the second term in (16) as:[70]

$$\log|B{*}| = 1/2\ \log|B{*}W\ B{*}'| - 1/2\ \log|W| \tag{29}$$

Using (28), we can evaluate the first term in (29)

----

[70] Noting that since B* and W are square matrices of the same order we have:
det(B* W)= det(B*).det(W)

$$B^*W\,B^{*\prime} = \begin{pmatrix} B_1^*W\,B^{*\prime} \\ \\ B_2^*W\,B^{*\prime} \end{pmatrix} = \begin{pmatrix} B_1^*W\,B^{*\prime} \\ \\ [(B^{*\prime})^{-1}]_2 B^{*\prime} \end{pmatrix}$$

$$= \begin{pmatrix} B_1^*W\,B_1^{*\prime} & B_1^*W\,B_2^{*\prime} \\ 0 & I_{22} \end{pmatrix} \tag{30}$$

since $B_1^* = B_1$, we have

$$\log|B^*W\,B^{*\prime}| = \log|B_1^*W\,B^{*\prime}| = \log|B_1W\,B_1'|$$

$$= \log(B_1W\,B_1') \tag{31}$$

Using (21), (29) and (31) we have the concentrated likelihood function in terms of A_1 and Σ_{11} as

$$L(A_1,\Sigma_{11}) = K + 1/2\,\log(B_1W\,B_1) - 1/2\,\log|W| - 1/2\,\log|\Sigma_{11}|$$

$$-1/2\,\mathrm{tr}(\Sigma_{11}^{-1}A_1M\,A_1') \tag{32}$$

If we partially differentiate (32) with respect to Σ_{11}(note that Σ_{11} is an scalar) we will get

$$-1/2\,(1/\,\Sigma_{11}) + 1/2(1/\,\Sigma_{11}^2)(A_1M\,A_1') = 0$$

therefore

$$\hat{\Sigma}_{11} = A_1M\,A_1' \tag{33}$$

Substituting $\hat{\Sigma}_{11}$ in (32), we will get

$$L(A_1) = K^* + 1/2\,\log(B_1W\,B_1') - 1/2\,\log|W|$$

$$-1/2\,\log(A_1M\,A_1') \tag{34}$$

Now write the first transformed structural equation as

$$B_1^*Y_1 + \Gamma_1^*X_1 - r\,B_1^*Y_{1,-1} - r\,\Gamma_1^*X_{1,-1} = \epsilon_1 \tag{35}$$

or

$$(B_1^* \quad \Psi_1^*)\begin{pmatrix} Y_1 \\ \Phi_1 \end{pmatrix} = \epsilon_1$$

and note that

$$A_1 = (B_1* \quad 0 \quad \Psi_1* \quad 0 \;)$$

therefore, the last term in (34) can be partitioned as

$$A_1 M A_1' = (B_1* \; \Psi_1*) M_1 (B_1* \; \Psi_1*)' = \epsilon_1 \epsilon_1' \tag{36}$$

where

$$M_1 = \begin{pmatrix} Y_1 Y_1' & Y_1 \Phi_1' \\ & \\ \Phi_1 Y_1' & \Phi_1 \Phi_1' \end{pmatrix}$$

Therefore, the concentrated likelihood function (34) becomes

$$L(B_1, \Gamma_1) = K* + 1/2 \log(B_1 W B_1') - 1/2 \log|W|$$

$$-1/2 \log(\epsilon_1 \epsilon_1') \tag{37}$$

Equation (37) is exactly the same as equation (14) of Amemiya(1966) except he considered the third term as part of the constant term.

However, the validity of (37) is on the basis of the validity of the partial derivatives given in (17) to (19). But in general (18) does not hold. This is because

$$\Psi_2* = [\Gamma*, (-RB)*, (-R\Gamma)*]_2 \tag{38}$$

Since the second term in Ψ_2* is a linear function of the elements of the second subset of the B matrix, the partial derivative (18) will be equal to

$$\frac{\partial \log B*}{\partial \Psi_{ik}*} = \begin{pmatrix} 0 \; \ldots \; 0 & \dfrac{\partial \log B*}{\partial (-RB)*} & 0 \ldots 0 \\ 0 \; \ldots \; 0 & & 0 \ldots 0 \end{pmatrix} \tag{39}$$

$$i = 2, \ldots, G$$

$$k = 1, \ldots, G+2K$$

(39) demonstrates that unless we ignore all the informations contain in Ψ_2*, we can not derive the Sargan and Amemiya

ALIML estimator. This is, in fact, the assumption made, but not explicitly explained, by Sargan(1961,p.421); this assumption does not affect the consistency of this estimator. However, its effect on the asymptotic efficiency of the Sargan ALIML is not addressed by Amemiya.

B. An Alternative Approach to the Derivation of LIML in the Presence of Autocorrelation

An alternative way of deriving the LIML in the presence of autocorrelation can be based on the restriction imposed on the structure of the R matrix. Instead of ignoring the information about the structure of the augmented reduced form equations other than the one under consideration, we can assume that the coefficients of autocorrelations are equal across equations. Under this assumption we can derive a LIML estimator which we shall refer to as the generalized LIML(GLIML). The derivation of GLIML is as follows. Consider system (1) again

$$B\ Y + \Gamma\ X = U \tag{40}$$

$$U = R\ U_{-1} + E$$

$$E(E_t E_t') = \Sigma$$

Assuming equal coefficients of autocorrelation across equations, we transform (40) with the aid of the (T-1).T Cochrane-Orcutt transformation matrix Q

$$Q = \begin{pmatrix} -r & 1 & 0 & 0 & \ldots & & \\ 0 & -r & 1 & 0 & 0 & & \ldots \\ 0 & 0 & 0 & 0 & -r & & 1 \end{pmatrix} \tag{41}$$

to the system free of autocorrelation

$$B\ Y\ Q' + \Gamma\ X\ Q' = U\ Q' = E$$

or

$$B\ \overline{Y} + \Gamma\ \overline{X} = E \tag{42}$$

or

$$A\ \overline{Z} = E$$

where

$$A = (\ B\ ,\ \Gamma)$$

$$\overline{Z}' = (\overline{Y}'\ ,\ \overline{X}')$$

The logarithmic likelihood function of the system (42) is

$$L = K_1 + T\ \log|B| - T/2\ \log|\Sigma| - 1/2\ \mathrm{tr}(\Sigma^{-1}AMA') \tag{43}$$

where

$$M = \overline{Z}\ \overline{Z}'$$

Since the estimation of the first equation of (40) is the focus of attention, we partition A and Σ such as

$$A = \begin{pmatrix} A_1 \\ \\ A_2 \end{pmatrix} = \begin{pmatrix} B_1 & \Gamma_1 \\ \\ B_2 & \Gamma_2 \end{pmatrix} \qquad \begin{matrix} A_1 \text{ is } 1.(G+K) \text{ matrix} \\ \\ A_2 \text{ is } (G-1)(G+K) \text{ matrix} \end{matrix}$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

A_1 and Σ_{11} are parameters to be estimated and A_2, $\Sigma_{12}=\Sigma_{21}'$ and Σ_{22} are parameters to be eliminated by partial maximization of (43). To do this we need to transform system (42) by a non-singular matrix such as F

$$F = \begin{pmatrix} 1 & 0 \\ F_{21} & F_{22} \end{pmatrix}$$

Transformation of the system (42) by F produces a new variance-covariance matrix such as

$$\Sigma * = F\ \Sigma\ F' = \begin{pmatrix} \Sigma_{11}* & \Sigma_{12}* \\ \Sigma_{21}* & \Sigma_{22}* \end{pmatrix} \tag{44}$$

where

$$\Sigma_{11}* = \Sigma_{11}$$

$$\Sigma_{21}*' = \Sigma_{12}* = \Sigma_{11}F_{21}' + \Sigma_{21}F_{22}' \tag{45}$$

$$\Sigma_{22}* = F_{21}\Sigma_{11}F_{21}' + F_{22}\Sigma_{21}F_{21}' + F_{21}\Sigma_{21}F_{22}' + F_{22}\Sigma_{22}F_{22}' \tag{46}$$

We will choose F_{21} so that making $\Sigma_{12}* = 0$. This means

$$F_{21}' = -\Sigma_{11}^{-1}\Sigma_{12}F_{22}' \tag{47}$$

substituting (47) into (46) yields

$$\Sigma_{22}* = F_{22}[\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}]F_{22}' \tag{48}$$

We will choose F_{22} so that making $\Sigma_{22}* = I_{22}$. Using the above results, the last term in (43) becomes

$$\Sigma^{-1}AMA' = \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & I_{22} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2* \end{pmatrix} M[A_1' \quad A_2*']$$

$$= \Sigma^{-1}A_1M\ A_1' + A_2*M\ A_2*' \tag{49}$$

Substituting (49) into (43) yields

$$1/T\ L* = K_2+\log|B*| -1/2\ \log|\Sigma_{11}| -1/2\ tr(\Sigma_{11}^{-1}A_1M\ A_1') - 1/2\ tr(A_2*M\ A_2*') \tag{50}$$

To obtain a concentrated likelihood function in terms of the

elements of the first equation, we have to eliminate A_2^* from (50) via a process partial maximization. Contribution to (50) from the elements of the second subset comes from the second and fifth terms. Therefore

$$\frac{\partial \log B^*}{\partial b_{ij}^*} = b^{*ji} = \frac{b_{ij}^{**}}{|B^*|} \qquad \begin{matrix} i=2,\dots,G \\ j=1,\dots,G \end{matrix} \tag{51}$$

where b_{ij}^*, b^{*ji} and b_{ij}^{**} are defined previously.

$$\partial \log|B^*| / \partial \Gamma_{ik}^* = 0 \;, \quad \begin{matrix} i=2,\dots,G \\ k=1,\dots,K \end{matrix} \tag{52}$$

Unlike equations (18), (52) holds unconditionally, since Γ_{ik}^* does not include any of the elements of B_{ij}^*.

$$\frac{\partial \operatorname{tr}(A_2^* M A_2^{*\prime})}{\partial A_2^*} = 2 A_2^* M \tag{53}$$

Combining (51) and (53) yields

$$\frac{\partial L^*}{\partial A_2^*} = \{[(B^{*\prime})^{-1}]_2 \quad 0\} - A_2^* M = 0 \tag{54}$$

Using the result in (54), the fifth term in (50) will become

$$-1/2 \operatorname{tr}(A_2^* M A_2^{*\prime}) = -1/2 \operatorname{tr}\{[(B^{*\prime})_2^{-1} \; 0] \begin{pmatrix} B_2^{*\prime} \\ \Gamma_2^{*\prime} \end{pmatrix}\}$$

$$= -1/2 \operatorname{tr}(I_{T-1}) = \text{const.} \tag{55}$$

To evaluate the second term in (50), we partition the last term in (54) as

$$A_2* \; M = [B_2* \quad \Gamma_2*]\begin{pmatrix} M_{\bar{y}\bar{y}} & M_{\bar{y}\bar{x}} \\ M_{\bar{x}\bar{y}} & M_{\bar{x}\bar{x}} \end{pmatrix} \tag{56}$$

where

$$\bar{Y} = YQ', \ldots$$

Therefore, we can write (54) as

$$\begin{aligned} [(B*')^{-1}]_2 &= B_2* \; M_{\bar{y}\bar{y}} + \Gamma_2* \; M_{\bar{x}\bar{y}} \\ 0 &= B_2* \; M_{\bar{y}\bar{x}} + \Gamma_2* \; M_{\bar{x}\bar{x}} \end{aligned} \tag{57}$$

The second equation in (57) gives

$$\Gamma_2* = -B_2 M_{yx} M_{xx}^{-1} \tag{58}$$

Substituting (58) into the first equation in (57) yields

$$[(B*')^{-1}]_2 = B_2*[M_{\bar{y}\bar{y}} - M_{\bar{y}\bar{x}} M_{\bar{x}\bar{x}}^{-1} M_{\bar{x}\bar{y}}] = B_2* \; \bar{W} \tag{59}$$

where

$$\bar{W} = \bar{Y}(I-\bar{X}'(\overline{XX}')^{-1}\bar{X})\bar{Y}'$$

Analogue to (29), the second term in (50) can be written as

$$\log|B*| = 1/2 \log|B*\bar{W} \; B*'| - 1/2 \log|\bar{W}| \tag{60}$$

Using (59), we can evaluate the first term in (60)

$$B*\bar{W} \; B*' = \begin{pmatrix} B_1*\bar{W} \; B*' \\ B_2*\bar{W} \; B*' \end{pmatrix} = \begin{pmatrix} B_1* \; \bar{W} \; B*' \\ [(B*')^{-1}]_2 B*' \end{pmatrix}$$

$$= \begin{pmatrix} B_1*\bar{W} \; B_1*' & B_1*\bar{W} \; B_2*' \\ 0 & I_{22} \end{pmatrix} \tag{61}$$

Note that $B_1* = B_1$, therefore

$$\log|B*\bar{W} \; B*'| = \log|B_1\bar{W} \; B_1'| = \log(B_1\bar{W} \; B_1') \tag{62}$$

Combining (50), (55), (60) and (62), the concentrated likelihood function interms of A_1 and Σ_{11} will become

$$L(A_1,\Sigma_{11}) = K + 1/2 \log(B_1\overline{W} B_1') - 1/2 \log|\overline{W}|$$
$$- 1/2 \log|\Sigma_{11}| - 1/2 \, tr(\Sigma_{11}^{-1} A_1 M A_1') \qquad (63)$$

Partial differentiation of (63) with respect to Σ_{11} yields

$$\hat{\Sigma}_{11} = A_1 M A_1' \qquad (64)$$

Substituting (64) in (63), we shall have

$$L(A_1) = K* + 1/2 \log(B_1\overline{W} B_1') - 1/2 \log|\overline{W}|$$
$$-1/2 \log(A_1 M A_1') \qquad (65)$$

Partitioning of the last term in (65) with respect to the coefficients of the first equation that are non-zero and those which are zero yields

$$A_1 = [B_1* \;\; 0 \;\; \Gamma_1* \;\; 0] \qquad (66)$$

$$A_1 M A_1' = (B_1* \;\; \Gamma_1*) M_1 (B_1* \;\; \Gamma_1*)' \qquad (67)$$

$$M_1 = \begin{pmatrix} \overline{Y}_1\overline{Y}_1' & \overline{Y}_1\overline{X}_1' \\ \overline{X}_1\overline{Y}_1' & \overline{X}_1\overline{X}_1' \end{pmatrix} \qquad (68)$$

Therefore

$$A_1 M A_1' = (B_1* \overline{Y}_1 + \Gamma_1* \overline{X}_1)(B_1* \overline{Y}_1 + \Gamma_1* \overline{X}_1)' \qquad (69)$$

Using (69) and noting that $B_1*=B_1$ and $\Gamma_1*=\Gamma_1$, the concentrated likelihood function (63) becomes

$$L(B_1*, \Gamma_1*) = K* + 1/2 \log(B_1\overline{W} B_1') - 1/2 \log|\overline{W}|$$
$$-1/2 \log(B_1 \overline{Y}_1 + \Gamma_1 \overline{X}_1)(B_1 \overline{Y}_1 + \Gamma_1 \overline{X}_1)' \qquad (70)$$

Note that B_1 is the first row of the B matrix. Maximization of (70) with respect to B_1 and Γ_1 for a given value of r yields the GLIML estimates of the parameters of the first equation. Note that we have to vary "r" between (-1,+1) and choose the estimates of B_1 and Γ_1 which maximizes (70). Derivation of an explicit formula for the

estimate of B_1 and Γ_1 along with the asymptotic distribution of GLIML are addressed in the chapter II.

C. Estimation of the Autoregressive System using all T observations

Extention of GLIML to the case using all T observations is straightforward. However, this is not true with Sargan and Amemiya's method. Due to the form of their transformation, their method cannot be extended to the case using all T observations.

Consider system (40) again

$$B\ Y + \Gamma\ X = U$$

$$U = R\ U_{-1} + E$$

Assuming equal coefficients of autocorrelation across equation, we can transform (71) with the aid of the Prais-Winsten transformation. The transformed system will be

$$B\ Y\ P' + \Gamma\ X\ P' = U\ P' = E \tag{72}$$

Derivation of the GLIML on the basis of the transformed variables in (72) is the same as the one for system (42) except for the additional term which is due to the Jacobian of the transformation. However, as we discussed earlier the omission of the Jacobian part of the likelihood function does not materially affect the performance of the estimator. Due to this fact and also the computational problems that inclusion of the Jacobian of the transformation creates, we have omitted it from the concentrated likelihood function associated with GLIML when it uses the Prais-Winsten

transformation.

APPENDIX II: COMPUTATIONAL STEPS OF DIFFERENT ESTIMATORS

In this appendix we shall first outline the computational steps and the computer programs used for the estimators designed for systems without lagged endogenous variables. Then the computational steps and the computer programs of the estimators for dynamic autoregresive systems will be outlined. In all the APL programs the following variables are global.

The exogenous variables which are

X_{11}, X_{22}, X_{33}, X_{44}, X_{55}, X_{66}

$Z = (X_{11}, X_{22}, X_{33}, X_{44}, X_{55}, X_{66}, C)$

where

C is a constant term which is a vector of ones.

YDATA is a matrix which contains all the observations on the endogenous variables.

A. Systems with no Lagged Endogenous Variables

1. Fair's Estimator

$y_1 = Y_1\beta_1 + X_1\gamma_1 + u_1 = Z_1\delta_1 + u_1$

1. $\Psi = (X_1 \; X_{1,-1} \; y_{1,-1} \; Y_{1,-1})$

Denote

$$\hat{Z}_1 = (\hat{Y}_1 \; X_1)$$

where

$$\hat{Y}_1 = \Psi(\Psi'\Psi)^{-1}\Psi'Y_1$$

$$\hat{\delta}_1 = (\hat{Z}_1'\hat{Z}_1)^{-1}\hat{Z}_1'y_1$$

2. $\hat{u}_1 = y_1 - Z_1\hat{\delta}_1$

$$\hat{r} = \sum_2^T \hat{u}_1'\hat{u}_{1,-1} / \sum_2^T \hat{u}_{1,-1}^2$$

3. Transforming the structural equation

$$\bar{y}_1 = y_1 - ry_{1,-1}$$

$$\bar{X}_1 = X_1 - rX_{1,-1}$$

$$\bar{Y}_1 = Y_1 - rY_{1,-1}$$

Denote

$$\hat{\bar{Z}}_1 = (\hat{\bar{Y}}_1 \; \bar{X}_1)$$

4. $\hat{\delta}_1 = (\hat{\bar{Z}}_1'\hat{\bar{Z}}_1)^{-1}\hat{\bar{Z}}_1'\bar{y}_1$

5. Iterate steps 2, 3, and 4 until convergence.

To calculate the variance-covariance matrix of δ_1

6. $\hat{\bar{u}}_1 = \bar{y}_1 - \bar{Z}_1\hat{\delta}_1$

$$\hat{\sigma}_{11} = \sum_1^T \hat{\bar{u}}_1'\hat{\bar{u}}_1/(T-K)$$

$$\mathrm{Var}(\hat{\delta}_1) = \hat{\sigma}_{11}(\hat{\bar{Z}}_1'\hat{\bar{Z}}_1)^{-1}$$

Note

$$K = (H - 1) + J$$

Where

H = Number of included endogenous variables.

J = Number of included exogenous variables.

```
     ∇ FAIR1
[ 1]   AFCOEF1← 0 7 ρ0
[ 2]   AVFCOEF← 0 5 ρ0
[ 3]   AFBAD1←0ρ0
[ 4]   II←1
[ 5]  L1:Y1←YDATA[(30×(II-1))+ι30;1]
[ 6]   Y2←YDATA[(30×(II-1))+ι30;2]
[ 7]   Y3←YDATA[(30×(II-1))+ι30;3]
[ 8]   IL1← ¯1 0 ↓Y1,Y2,Y3,X11,X44,C
[ 9]   ZR←(1 0 ↓X11,X44),IL1
[10]   Y2H←ZR+.×(1↓Y2)⌹ZR
[11]   Y3H←ZR+.×(1↓Y3)⌹ZR
[12]   B1←(1↓Y1)⌹(Y2H,Y3H, 1 0 ↓X1)
[13]   R←I←0
[14]  L2:R1←R
[15]   R←(1↓E)⌹¯1↓E←Y1-(Y2,Y3,X1)+.×B1
[16]   AFBAD1←AFBAD1,II×ι(R≥1)
[17]   R←(R,0.99)[1+R≥1]
[18]   X1R←(1 0 ↓X1)-R× ¯1 0 ↓X1
[19]   Y1R←(1↓Y1)-R×¯1↓Y1
[20]   Y2R←(1↓Y2)-R×¯1↓Y2
[21]   Y3R←(1↓Y3)-R×¯1↓Y3
[22]   Y2HR←Y2H-R×¯1↓Y2
[23]   Y3HR←Y3H-R×¯1↓Y3
[24]   B1←Y1R⌹(Y2HR,Y3HR,X1R)
[25]   →L2×ι(20≥I←I+1)∧(0.005≤|R-R1)
[26]   AFCOEF1←AFCOEF1,[1](B1,R,I)
[27]   URH←Y1R-(Y2R,Y3R,X1R)+.×B1
[28]   S2←(+/URH*2)÷26
[29]   VAR←S2×⌹(⍉(Y2HR,Y3HR,X1R))+.×Y2HR,Y3HR,X1R
[30]   AVFCOEF←AVFCOEF,[1] VAR[1;1],VAR[2;2],VAR[3;3],VAR[4;4
       ],VAR[5;5]
[31]   →L1×ι(100≥II←II+1)
[32]   '***END'
     ∇
```

2. **Fair-Brundy and Jorgenson Estimator**

$y_1 = Y_1 B_1 + X_1 \gamma_1 + u_1 = Z_1 \delta_1 + u_1$

The augmented reduced form is

$$Y_t = -X_t \Gamma B^{-1} + Y_{t-1} BRB^{-1} + X_{t-1} \Gamma RB^{-1} + E_t B^{-1}$$

or

$$Y_t = X_t \Pi_1 + Y_{t-1} \Pi_2 + X_{t-1} \Pi_3 + V$$

1. Estimate all structural equations by 2SLS and use the estimated coefficient to form a consistent estimate of Π_1, Π_2, and Π_3. Also from the estimated residuals obtain estimates of the autocorrelation coefficients and form the R matrix. Then obtain the fitted value for Y_1 as

$$\hat{Y}_1 = X_t \hat{\Pi}_{11} + Y_{t-1} \hat{\Pi}_{21} + X_{t-1} \hat{\Pi}_{31}$$

where

$$\hat{\Pi}_1 = \hat{\Gamma}\ \hat{B}^{-1}$$

$$\hat{\Pi}_2 = \hat{B}\ \hat{R}\ \hat{B}^{-1}$$

$$\hat{\Pi}_3 = \hat{\Gamma}\ \hat{R}\ \hat{B}^{-1}$$

2. $\hat{u}_1 = y_1 - Z_1 \hat{\delta}_1$

where $\hat{\delta}_1$ is estimated in the first step.

$$\hat{r} = \sum_{2}^{T} \hat{u}_1' \hat{u}_{1,-1} / \sum_{2}^{T} \hat{u}_{1,-1}^2$$

3. Use $\hat{r}$, transform the first structural equation as

$$y_1 - \hat{r} y_{1,-1} = (Y_1 - \hat{r} Y_{1,-1}) B_1 + (X_1 - \hat{r} X_{1,-1}) \gamma_1 + \epsilon_1$$

Define

$$\bar{Z}_1 = [(Y_1 - \hat{r} Y_{1,-1})\ \ (X_1 - \hat{r} X_{1,-1})].$$

$$\bar{y}_1 = y_1 - \hat{r} y_{1,-1}$$

$$W_1 = [(\hat{Y}_1 - \hat{r} Y_{1,-1})\ \ (X_1 - \hat{r} X_{1,-1})]$$

4. $\hat{\delta}_1 = (W_1' Z_1)^{-1} W_1' \bar{y}_1$

5. Iterate steps 2, 3, and 4 until convergence.

To calculate the variance-covariance matrix of δ

6. $\hat{\bar{u}}_1 = \bar{y}_1 - \bar{Z}_1\hat{\delta}_1$

$$\hat{\sigma}_{11} = \Sigma\hat{\bar{u}}_1{}'\hat{\bar{u}}_1/(T-K)$$

$$\mathrm{Var}(\hat{\delta}_1) = \sigma_{11}(W_1{}'\bar{Z}_1)^{-1}$$

```
    ∇ FBRUNDY
[ 1]   II←1
[ 2]   AFBBAD←0⍴0
[ 3]   AFBCOEF← 0 7 ⍴0
[ 4]   AVFBR← 0 5 ⍴0
[ 5]  LL1:Y1←YDATA[(30×(II-1))+⍳30;1]
[ 6]   Y2←YDATA[(30×(II-1))+⍳30;2]
[ 7]   Y3←YDATA[(30×(II-1))+⍳30;3]
[ 8]   Y1H←Z+.×Y1⌹Z
[ 9]   Y2H←Z+.×Y2⌹Z
[10]   Y3H←Z+.×Y3⌹Z
[11]   B1←Y1⌹(Y2H,Y3H,X1)
[12]   B2←Y2⌹(Y1H,X2)
[13]   B3←Y3⌹(Y2H,X3)
[14]   R1←(1↓E)⌹¯1↓E←Y1-(Y2,Y3,X1)+.×B1
[15]   R2←(1↓E)⌹¯1↓E←Y2-(Y1,X2)+.×B2
[16]   R3←(1↓E)⌹¯1↓E←Y3-(Y2,X3)+.×B3
[17]   B11←-B1[1]
[18]   B12←-B1[2]
[19]   B21←-B2[1]
[20]   B31←-B3[1]
[21]   R← 3 3 ⍴R1,0,0,0,R2,0,0,0,R3
[22]   BB← 3 3 ⍴1,B11,B12,B21,1,0,0,B31,1
[23]   Q1← 1 7 ⍴B1[3],0,0,B1[4],0,0,B1[5]
[24]   Q2← 1 7 ⍴0,B2[2],0,B2[3],0,B2[4],B2[5]
[25]   Q3← 1 7 ⍴0,B3[2],B3[3],0,B3[4],0,B3[5]
[26]   Q←Q1,[1] Q2,[1] Q3
[27]   P1←(⌹BB)+.×R+.×BB
[28]   P2←(⌹BB)+.×Q
[29]   P3←(⌹BB)+.×R+.×Q
[30]   YY←⍉(3 30 ⍴(Y1,Y2,Y3))
[31]   Y2H←(P1[2;]+.×⍉(¯1 0 ↓YY))+(P2[2;]+.×⍉(1 0 ↓Z))-P3[2;]
       +.×⍉ ¯1 0 ↓Z
[32]   Y3H←(P1[3;]+.×⍉(¯1 0 ↓YY))+(P2[3;]+.×⍉(1 0 ↓Z))-P3[3;]
       +.×⍉ ¯1 0 ↓Z
[33]   I←0
[34]  L1:RR←R1
[35]   X1R←(1 0 ↓X1)-R1× ¯1 0 ↓X1
[36]   Y1R←(1↓Y1)-R1×¯1↓Y1
[37]   Y2HR←Y2H-R1×¯1↓Y2
[38]   Y3HR←Y3H-R1×¯1↓Y3
[39]   Y2R←(1↓Y2)-R1×¯1↓Y2
[40]   Y3R←(1↓Y3)-R1×¯1↓Y3
[41]   W←Y2HR,Y3HR,X1R
[42]   M←Y2R,Y3R,X1R
[43]   BB1←(⌹((⍉W)+.×M))+.×((⍉W)+.×Y1R)
[44]   R1←(1↓E)⌹¯1↓E←Y1-(Y2,Y3,X1)+.×BB1
[45]   AFBBAD←AFBBAD,II×⍳(R1≥1)
[46]   R1←(R1,0.99)[1+R1≥1]
[47]   →L1×⍳(20≥I←I+1)∧(0.005≤|RR-R1)
[48]   AFBCOEF←AFBCOEF,[1](BB1,R1,I)
[49]   UH←Y1R-(Y2R,Y3R,X1R)+.×BB1
[50]   S2←(+/UH⋆2)÷26
[51]   VAR←S2×⌹((⍉W)+.×M)
```

```
[52]  AVFBR←AVFBR,[1] VAR[1;1],VAR[2;2],VAR[3;3],VAR[4;4],VA
      R[5;5]
[53]  →LL1×⍳(1≥II←II+1)
[54]  '***END'
    ∇
```

3. Modified Brundy and Jorgenson Estimator(T Observation)

$y_1 = Y_1 B_1 + X_1 \delta_1 + u_1 = Z_1 \delta_1 + u_1$

Using the ordinary reduced form

$$Y = -X \Gamma B^{-1} + U B^{-1} = X \Pi_1 + V$$

1. Run 2SLS on all the structural equations and use the estimated coefficients to form a consistent estimate of Π_1. Then obtain

$$\hat{Y}_1 = X \hat{\Pi}_{11}$$

2. $\hat{u}_1 = y_1 - Z_1 \hat{\delta}_1$

$$\hat{r} = \sum_{2}^{T} \hat{u}_1{}' \hat{u}_{1,-1} / \sum_{3}^{T} \hat{u}_{1,-1}^2 \quad \text{(Prais-Winsten formula)}$$

3. Transform the first structural equation using the Prais-Winsten transformation matrix and denote

$$\dot{Z}_1 = (\hat{P}\, Y_1 \quad \hat{P}\, X_1)$$

$$W_1 = (\hat{P}\, \hat{Y}_1 \quad \hat{P}\, X_1)$$

Where $\hat{P}$ is constructed using estimated autocorrelation coefficient in the second step.

4. $\hat{\delta}_1 = (W_1{}' \dot{Z}_1)^{-1} W_1{}' (\hat{P} y_1)$

5. Iterate steps 2, 3, and 4 until convergence.

Calculation of variance-covariance matrix

6. $\dot{\hat{u}}_1 = \hat{p}\, y_1 - (\hat{P}\, Y_1 \quad \hat{P}\, X_1)\hat{\delta}_1$

$$\hat{\sigma}_{11} = \Sigma \dot{\hat{u}}_1{}' \dot{\hat{u}}_1 / (T-K)$$

$$\mathrm{Var}(\hat{\delta}_1) = \sigma_{11} (W_1{}' \dot{Z}_1)^{-1}$$

```
    ∇ BRUNDY
[ 1]  II←1
[ 2]  ABBCOEF← 0 7 ⍴0
[ 3]  ABBAD←0⍴0
[ 4]  AVBR← 0 5 ⍴0
[ 5] LL1:Y1←YDATA[(30×(II-1))+⍳30;1]
[ 6]  Y2←YDATA[(30×(II-1))+⍳30;2]
[ 7]  Y3←YDATA[(30×(II-1))+⍳30;3]
[ 8]  Y1H←Z+.×Y1⌹Z
[ 9]  Y2H←Z+.×Y2⌹Z
[10]  Y3H←Z+.×Y3⌹Z
[11]  B1←Y1⌹(Y2H,Y3H,X1)
[12]  B2←Y2⌹(Y1H,X2)
[13]  B3←Y3⌹(Y2H,X3)
[14]  B11←-B1[1]
[15]  B12←-B1[2]
[16]  B21←-B2[1]
[17]  B31←-B3[1]
[18]  BB← 3 3 ⍴1,B11,B12,B21,1,0,0,B31,1
[19]  Q1← 1 7 ⍴B1[3],0,0,B1[4],0,0,B1[5]
[20]  Q2← 1 7 ⍴0,B2[2],0,B2[3],0,B2[4],B2[5]
[21]  Q3← 1 7 ⍴0,B3[2],B3[3],0,B3[4],0,B3[5]
[22]  Q←Q1,[1] Q2,[1] Q3
[23]  P←(⌹BB)+.×Q
[24]  Y2H←P[2;]+.×(⌽Z)
[25]  Y3H←P[3;]+.×(⌽Z)
[26]  R←R1←I←0
[27] L1:R1←R
[28]  R←(1↓E)⌹¯1↓E←Y1-(Y2,Y3,X1)+.×B1
[29]  ABBAD←ABBAD,II×⍳(R≥1)
[30]  R←(R,0.99)[1+R≥1]
[31]  X1R←(1 0 ↓X1)-R× ¯1 0 ↓X1
[32]  Y1R←(1↓Y1)-R×¯1↓Y1
[33]  Y2HR←(1↓Y2H)-R×¯1↓Y2H
[34]  Y3HR←(1↓Y3H)-R×¯1↓Y3H
[35]  Y2R←(1↓Y2)-R×¯1↓Y2
[36]  Y3R←(1↓Y3)-R×¯1↓Y3
[37]  W←Y2HR,Y3HR,X1R
[38]  M←Y2R,Y3R,X1R
[39]  B1←(⌹((⌽W)+.×M))+.×((⌽W)+.×Y1R)
[40]  →L1×⍳(20≥I←I+1)∧(0.005≤|R-R1)
[41]  ABBCOEF←ABBCOEF,[1](B1,R,I)
[42]  UH←Y1R-(Y2R,Y3R,X1R)+.×B1
[43]  S2←(+/UH⋆2)÷26
[44]  VAR←S2×⌹((⌽W)+.×M)
[45]  AVBR←AVBR,[1] VAR[1;1],VAR[2;2],VAR[3;3],VAR[4;4],VAR[
      5;5]
[46]  →LL1×⍳(100≥II←II+1)
[47]  '***END'
    ∇
```

4. Modified 2SLS

$$y_1 = Y_1 B_1 + X_1\gamma_1 + u_1 = Z_1\delta_1 + u_1$$

1. $\hat{Y}_1 = X(X'X)^{-1}X'Y_1$

Denote

$$\hat{Z}_1 = (\hat{Y}_1 \ X_1)$$

$$\hat{\delta}_1 = (\hat{Z}_1'\hat{Z}_1)^{-1}\hat{Z}_1'y_1$$

2. $\hat{u}_1 = y_1 - Z_1\hat{\delta}_1$

$$\hat{r} = \sum_2^T \hat{u}_1'\hat{u}_{1,-1} / \sum_2^T \hat{u}_{1,-1}^2$$

3. Transforming the structural equation as

$$\bar{y}_1 = y_1 - \hat{r}\, y_{1,-1}$$

$$\bar{X}_1 = X_1 - \hat{r}\, X_{1,-1}$$

$$\bar{Y}_1 = Y_1 - \hat{r}\, Y_{1,-1}$$

$$\hat{\bar{Y}}_1 = \bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}'\bar{Y}_1$$

where $\bar{X} = Q\,X$, and Q is the CORC matrix.

4. Denote

$$\hat{\bar{Z}}_1 = (\hat{\bar{Y}}_1 \ \ \bar{X}_1)$$

Therefore

$$\hat{\delta}_1 = (\hat{\bar{Z}}_1'\hat{\bar{Z}}_1)^{-1}\hat{\bar{Z}}_1'\bar{y}_1$$

5. Iterate steps 2, 3, and 4 until convergence.

Calculation of the variance-covariance matrix of δ

6. $\hat{\bar{u}}_1 = \bar{y}_1 - (\bar{Y}_1 \ \bar{X}_1)\hat{\delta}_1$

$$\hat{\sigma}_{11} = \sum_2^T \hat{\bar{u}}_1'\hat{\bar{u}}_1/(T-K)$$

$$\mathrm{Var}(\hat{\delta}_1) = \hat{\sigma}_{11}(\hat{\bar{Z}}_1'\hat{\bar{Z}}_1)^{-1}$$

```
    ∇ TSLS
[ 1]  ATSCOEF← 0 7 ρ0
[ 2]  ATSBAD←0ρ0
[ 3]  AVTSLS← 0 5 ρ0
[ 4]  II←1
[ 5] LL1:Y1←YDATA[(30×(II-1))+ι30;1]
[ 6]  Y2←YDATA[(30×(II-1))+ι30;2]
[ 7]  Y3←YDATA[(30×(II-1))+ι30;3]
[ 8]  Y2H←Z+.×Y2⌹Z
[ 9]  Y3H←Z+.×Y3⌹Z
[10]  B1←Y1⌹(Y2H,Y3H,X1)
[11]  R←I←0
[12] L1:R1←R
[13]  R←(1↓E)⌹¯1↓E←Y1-(Y2,Y3,X1)+.×B1
[14]  TSBAD←TSBAD,II×ι(R≥1)
[15]  R←(R,0.99)[1+R≥1]
[16]  Y1R←(1↓Y1)-R×¯1↓Y1
[17]  Y2R←(1↓Y2)-R×¯1↓Y2
[18]  Y3R←(1↓Y3)-R×¯1↓Y3
[19]  ZR←(1 0 ↓Z)-R× ¯1 0 ↓Z
[20]  Y2HR←ZR+.×Y2R⌹ZR
[21]  Y3HR←ZR+.×Y3R⌹ZR
[22]  X1R←(1 0 ↓X1)-R× ¯1 0 ↓X1
[23]  B1←Y1R⌹(Y2HR,Y3HR,X1R)
[24]  →L1×ι(20≥I←I+1)∧(0.005≤|R-R1)
[25]  ATSCOEF←ATSCOEF,[1](B1,R,I)
[26]  UH←Y1R-(Y2R,Y3R,X1R)+.×B1
[27]  S2←(+/UH*2)÷26
[28]  VAR←S2×⌹(⍉(Y2HR,Y3HR,X1R))+.×Y2HR,Y3HR,X1R
[29]  AVTSLS←AVTSLS,[1] VAR[1;1],VAR[2;2],VAR[3;3],VAR[4;4],
      VAR[5;5]
[30]  →LL1×ι(100≥II←II+1)
[31]  '***END'
    ∇
```

5. Theil G2SLS

$y_1 = Y_1 B_1 + X_1 \gamma_1 + u_1 = Z_1 \delta_1 + u_1$

1. $\hat{Y}_1 = X \hat{\Pi}_1 = X(X'X)^{-1} X' Y_1$

Denote

$$\hat{Z}_1 = (\hat{Y}_1 \; X_1)$$

$$\hat{\delta}_1 = (\hat{Z}_1' \hat{Z}_1)^{-1} \hat{Z}_1' y_1$$

2. $\hat{u}_1 = y_1 - Z_1 \hat{\delta}_1$

$$\hat{r} = \sum_{2}^{T} \hat{u}_1' \hat{u}_{1,-1} / \sum_{3}^{T} \hat{u}_{1,-1}^2 \qquad \text{(Prais-Winsten)}$$

3. Transforming the first equation with the Prais-Winsten transformation matrix P

$$\dot{y}_1 = \dot{Y}_1 B_1 + \dot{X}_1 \gamma_1 + \epsilon_1$$

$$\hat{\dot{Y}}_1 = \dot{X} \hat{\Pi}_1 = \dot{X}(\dot{X}'\dot{X})^{-1} \dot{X}' \dot{Y}_1$$

Denote

$$\hat{\dot{Z}}_1 = (\hat{\dot{Y}}_1 \;\; \dot{X}_1)$$

4. $\hat{\delta}_1 = (\hat{\dot{Z}}_1' \hat{\dot{Z}}_1)^{-1} \hat{\dot{Z}}_1' \dot{y}_1$

5. Iterate steps 2, 3, and 4 until convergence.

Calculation of the variance-covariance matrix

6. $\hat{\dot{u}}_1 = \dot{y}_1 - (\dot{Y}_1 \;\; \dot{X}_1) \hat{\delta}_1$

$$\hat{\sigma}_{11} = \Sigma \hat{\dot{u}}_1' \hat{\dot{u}}_1 / (T-K)$$

$$\mathrm{Var}(\hat{\delta}_1) = \hat{\sigma}_{11} (\hat{\dot{Z}}_1' \hat{\dot{Z}}_1)^{-1}$$

```
    ∇ THEIL1
[ 1]  C←30ρ1
[ 2]  ATCOEF1← 0 7 ρ0
[ 3]  ATBAD1←0ρ0
[ 4]  AVTHEIL← 0 5 ρ0
[ 5]  II←1
[ 6] LL1:Y1←YDATA[(30×(II-1))+⍳30;1]
[ 7]  Y2←YDATA[(30×(II-1))+⍳30;2]
[ 8]  Y3←YDATA[(30×(II-1))+⍳30;3]
[ 9]  Y2H←Z+.×Y2⌹Z
[10]  Y3H←Z+.×Y3⌹Z
[11]  B1←Y1⌹(Y2H,Y3H,X1)
[12]  R1←I←0
[13] L1:R11←R1
[14]  R1←((1↓E)+.×¯1↓E)÷+/(1↓¯1↓E←Y1-(Y2,Y3,X1)+.×B1)*2
[15]  ATBAD1←ATBAD1,II×⍳(R1≥1)
[16]  R1←(R1,0.99)[1+R1≥1]
[17]  RR1←(1-R1*2)*0.5
[18]  ZR← 1 7 ρZ[1;]×RR1
[19]  ZR←ZR,[1](1 0 ↓Z)-R1× ¯1 0 ↓Z
[20]  X1R← 1 3 ρX1[1;]×RR1
[21]  X1R←X1R,[1](1 0 ↓X1)-R1× ¯1 0 ↓X1
[22]  Y1R←Y1[1]×RR1
[23]  Y1R←Y1R,(1↓Y1)-R1×¯1↓Y1
[24]  Y2R←Y2[1]×RR1
[25]  Y2R←Y2R,(1↓Y2)-R1×¯1↓Y2
[26]  Y3R←Y3[1]×RR1
[27]  Y3R←Y3R,(1↓Y3)-R1×¯1↓Y3
[28]  Y2HH←ZR+.×Y2R⌹ZR
[29]  Y3HH←ZR+.×Y3R⌹ZR
[30]  B1←Y1R⌹(Y2HH,Y3HH,X1R)
[31]  →L1×⍳(20≥I←I+1)∧(0.005≤|R1-R11)
[32]  ATCOEF1←ATCOEF1,[1](B1,R1,I)
[33]  UH←Y1R-(Y2R,Y3R,X1R)+.×B1
[34]  S2←(+/UH*2)÷26
[35]  VAR←S2×⌹(⍉(Y2HH,Y3HH,X1R))+.×Y2HH,Y3HH,X1R
[36]  AVTHEIL←AVTHEIL,[1] VAR[1;1],VAR[2;2],VAR[3;3],VAR[4;4
      ],VAR[5;5]
[37]  →LL1×⍳(100≥II←II+1)
[38]  'END OF THE RUN'
    ∇
```

6. **Theil: Instrumental Variable Approach**

$y_1 = Y_1B_1 + X_1\gamma_1 + u_1 = Z_1\delta_1 + u_1$

1. $\hat{Y}_1 = X\hat{\Pi}_1 = X(X'X)^{-1}X'Y_1$

Denote

$$\hat{Z}_1 = (\hat{Y}_1 \; X_1)$$

$$\hat{\delta}_1 = (\hat{Z}_1'\hat{Z}_1)^{-1}\hat{Z}_1'y_1$$

2. $\hat{u}_1 = y_1 - Z_1\hat{\delta}_1$

$$\hat{r} = \sum_{2}^{T}\hat{u}_1'\hat{u}_{1,-1} / \sum_{3}^{T}\hat{u}_{1,-1}^2 \qquad \text{(Prais-Winsten)}$$

3. Transforming the first structural equation using Prais-Winsten transformation matrix

$$Py_1 = PY_1B_1 + PX_1\gamma_1 + Pu_1$$

or

$$\dot{y}_1 = \dot{Y}_1B_1 + \dot{X}_1\gamma_1 + \epsilon_1 = \dot{Z}_1\delta_1 + \epsilon_1$$

4. Denote

$$\hat{W}_1 = (\dot{\hat{Y}}_1 \;\; \dot{X}_1)$$

where

$$\hat{Y}_1 = X(X'X)^{-1}X'Y_1$$

$$\dot{\hat{Y}}_1 = P\hat{Y}_1$$

5. $\hat{\delta}_1 = (\hat{W}_1'\dot{Z}_1)^{-1}\hat{W}_1'\dot{y}_1$

6. Iterate steps 2, 3, 4, and 5 until convergence.

Calculation of the variance-covariance matrix

7. $\dot{\hat{u}}_1 = \dot{y}_1 - \dot{Z}_1\hat{\delta}_1$

$$\hat{\sigma}_{11} = \Sigma\dot{\hat{u}}_1'\dot{\hat{u}}_1/(T-K)$$

$$\mathrm{Var}(\hat{\delta}_1) = \hat{\sigma}_{11}(\hat{W}_1'\dot{Z}_1)^{-1}$$

```
    ∇ THEIL1IV
[ 1]   ATV1← 0 7 ρ0
[ 2]   AVTV1← 0 5 ρ0
[ 3]   ATVBAD1←0ρ0
[ 4]   II←1
[ 5]  LL1:Y1←YDATA[(30×(II-1))+ι30;1]
[ 6]   Y2←YDATA[(30×(II-1))+ι30;2]
[ 7]   Y3←YDATA[(30×(II-1))+ι30;3]
[ 8]   Y2H←Z+.×Y2⌹Z
[ 9]   Y3H←Z+.×Y3⌹Z
[10]   B1←Y1⌹(Y2H,Y3H,X1)
[11]   R←I←0
[12]  L1:R1←R
[13]   R←((1↓E)+.×¯1↓E)÷+/(1↓¯1↓E←Y1-(Y2,Y3,X1)+.×B1)*2
[14]   ATVBAD1←ATVBAD1,II×ι(R≥1)
[15]   R←(R,0.99)[1+R≥1]
[16]   RR←(1-R*2)*0.5
[17]   X1R← 1 3 ρX1[1;]×RR
[18]   X1R←X1R,[1](1 0 ↓X1)-R× ¯1 0 ↓X1
[19]   Y1R←Y1[1]×RR
[20]   Y1R←Y1R,(1↓Y1)-R×¯1↓Y1
[21]   Y2R←Y2[1]×RR
[22]   Y2R←Y2R,(1↓Y2)-R×¯1↓Y2
[23]   Y3R←Y3[1]×RR
[24]   Y3R←Y3R,(1↓Y3)-R×¯1↓Y3
[25]   Y2HR←Y2H[1]×RR
[26]   Y2HR←Y2HR,(1↓Y2H)-R×¯1↓Y2H
[27]   Y3HR←Y3H[1]×RR
[28]   Y3HR←Y3HR,(1↓Y3H)-R×¯1↓Y3H
[29]   W←Y2HR,Y3HR,X1R
[30]   M←Y2R,Y3R,X1R
[31]   B1←(⌹((⍉W)+.×M))+.×((⍉W)+.×Y1R)
[32]   →L1×ι(20≥I←I+1)∧(0.005≤|R-R1)
[33]   ATV1←ATV1,[1](B1,R,I)
[34]  ⍝ ***END OF THE FIRST EQUATION
[35]   UH←Y1R-(Y2R,Y3R,X1R)+.×B1
[36]   S2←(+/UH*2)÷26
[37]   VAR←S2×⌹(⍉W)+.×M
[38]   AVTV1←AVTV1,[1] VAR[1;1],VAR[2;2],VAR[3;3],VAR[4;4],VA
       R[5;5]
[39]   →LL1×ι(100≥II←II+1)
[40]   '****END OF THE RUN'
    ∇
```

7. Amemiya LIML (ALIML)

$y_1 = Y_1B_1 + X_1\gamma_1 + u_1 = Z_1\delta_1 + u_1$

1. Run 2SLS on the first structural equation and use the estimated coefficients to calculate the autocorrelation coefficient as

2. $\hat{u}_1 = y_1 - Z_1\hat{\delta}_1$

$$\hat{r} = \sum_{2}^{T}\hat{u}_1{}'\hat{u}_{1,-1} / \sum_{2}^{T}\hat{u}_{1,-1}{}^2$$

3. Write the structural equation under consideration as

$$y_1{*}B_1{*} = X_1\gamma_1 + u_1$$

where

$$y_1{*} = (y_1 \quad Y_1)$$

$$B_1{*} = (1 \quad -B_1')'$$

4. Denote

$$\Psi_1 = \overline{y}_1{*}'(I-\overline{X}_1(\overline{X}_1'\overline{X}_1)^{-1}\overline{X}_1')\overline{y}_1{*}$$

where the variables are transformed by CORC matrix. Denote

$$\overline{y}_1{*} = (\overline{y}_1 \quad \overline{Y}_1)$$

$$\Psi_2 = y_1{*}'(I-\Phi(\Phi'\Phi)^{-1}\Phi')y_1{*}$$

where

$$\Phi = (X \; Y_{-1} \; X_{-1})$$

$$X = (X_1 \quad X^{x})$$

$$y = (y_1{*} \quad y^{x})$$

where X^x and y^x are the exogenous and endogenous variables excluded from the first equation.

5. Find λ as the minimum root of

$$\det(\Psi_1 - \lambda\Psi_2) = 0$$

6. Estimate $B_1{*}$ as the solution to the following equation

$$(\Psi_1 - \hat{\lambda}_{min}\Psi_2)\ B_2* = 0$$

Use $\hat{B}_1*$ to estimate γ_1 as

$$\hat{\gamma}_1 = (\bar{X}_1'\bar{X}_1)^{-1}\bar{X}_1'\bar{y}_1*\hat{B}_1*$$

7. Iterate steps 2, 3, 4, 5, and 6 until convergence.

Calculation of the variance-covariance matrix of δ_1

8. $\hat{\bar{u}}_1 = \bar{y}_1 - (\bar{Y}_1\ \bar{X}_1)\hat{\delta}_1$

$$\hat{\sigma}_{11} = \Sigma\hat{\bar{u}}_1'\hat{\bar{u}}_1/(T-K)$$

$$\mathrm{Var}(\hat{\delta}_1) = \hat{\sigma}_{11}[(\hat{\bar{Y}}_1\ \bar{X}_1)'(\hat{\bar{Y}}_1\ \bar{X}_1)]^{-1}$$

Where $\hat{\bar{Y}}_1 = \Phi(\Phi'\Phi)^{-1}\Phi'\bar{Y}_1$

```
    ∇ ALIML
[ 1]  AAC1← 0 7 ρ0
[ 2]  AVAC1← 0 5 ρ0
[ 3]  AABAD←0ρ0
[ 4]  II←1
[ 5] LL1:Y1←YDATA[(30×(II-1))+⍳30;1]
[ 6]  Y2←YDATA[(30×(II-1))+⍳30;2]
[ 7]  Y3←YDATA[(30×(II-1))+⍳30;3]
[ 8]  Y2H←Z+.×Y2⌹Z
[ 9]  Y3H←Z+.×Y3⌹Z
[10]  B1←Y1⌹(Y2H,Y3H,X1)
[11]  YL← ¯1 0 ↓⍉(3 30 ρ(Y1,Y2,Y3))
[12]  YA← 1 0 ↓⍉(3 30 ρ(Y1,Y2,Y3))
[13]  V←(1 0 ↓X11,X22,X33,X44,X55,X66),YL, ¯1 0 ↓XDATA,C
[14]  R←I←0
[15] L1:R1←R
[16]  R←(1↓E)⌹¯1↓E←Y1-(Y2,Y3,X1)+.×B1
[17]  AABAD←AABAD,II×⍳(R≥1)
[18]  R←(R,0.99)[1+R≥1]
[19]  X1R←(1 0 ↓X1)-R× ¯1 0 ↓X1
[20]  Y1R←(1↓Y1)-R×¯1↓Y1
[21]  Y2R←(1↓Y2)-R×¯1↓Y2
[22]  Y3R←(1↓Y3)-R×¯1↓Y3
[23]  YY←⍉(3 29 ρ(Y1R,Y2R,Y3R))
[24]  W1←((⍉YY)+.×YY)-((⍉YY)+.×X1R)+.×YY⌹X1R
[25]  W2←((⍉YA)+.×YA)-((⍉YA)+.×V)+.×YA⌹V
[26]  L←1+÷MXEV Q←(⌹(W1-W2))+.×W2
[27]  WW←W1-L×W2
[28]  BN←(-WW[;1])⌹ 0 1 ↓WW
[29]  B1←(-BN),(⌹(⍉X1R)+.×X1R)+.×(⍉X1R)+.×YY+.×1,BN
[30]  →L1×⍳(20≥I←I+1)∧(0.005≤|R-R1)
[31]  AAC1←AAC1,[1](B1,R,I)
[32]  UH←Y1R-(Y2R,Y3R,X1R)+.×B1
[33]  S2←(+/UH*2)÷26
[34]  Y2HR←V+.×Y2R⌹V
[35]  Y3HR←V+.×Y3R⌹V
[36]  VAR←S2×⌹(⍉(Y2HR,Y3HR,X1R))+.×Y2HR,Y3HR,X1R
[37]  AVAC1←AVAC1,[1] VAR[1;1],VAR[2;2],VAR[3;3],VAR[4;4],VA
      R[5;5]
[38]  →LL1×⍳(100≥II←II+1)
[39]  '***END'
    ∇
```

8. **Generalized LIML (GLIML)**

$y_1 = Y_1 B_1 + X_1\gamma_1 + u_1 = Z_1\delta_1 + u_1$

1. Write the structural equation under consideration as

$$y_1{*}B_1{*} = X_1\gamma_1 + u_1$$

Where y_1* and B_1* are defined above.

2. For given r transform the above structural equation, using P.W matrix, to the one free of autocorrelation

$$\dot{y}_1{*}B_1{*} = \dot{X}_1\gamma_1 + \epsilon_1$$

3. Denote

$$\Psi_1{*} = \dot{y}_1{*}'(I-\dot{X}_1(\dot{X}_1'\dot{X}_1)^{-1}\dot{X}_1')\dot{y}_1{*}$$

$$\Psi_2{*} = \dot{y}_1{*}'(I-\dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}')\dot{y}_1{*}$$

Calculate λ as the minimum root of

$$\det(\Psi_1{*} - \lambda\Psi_2{*}) = 0$$

4. Estimate B_1* and γ_1 as the solutions to the equations

$$(\Psi_1{*} - \lambda\Psi_2{*})\hat{B}_1{*} = 0$$

$$\hat{\gamma}_1 = (\dot{X}_1'\dot{X}_1)^{-1}\dot{X}_1'\dot{y}_1{*}\hat{B}_1{*}$$

5. $\hat{\dot{u}}_1 = \dot{y}_1 - (\dot{Y}_1 \quad \dot{X}_1)\hat{\delta}_1$

6. Calculate steps 1, 2, 3, 4, and 5 for r varying between (-1,+1) and choose the estimated coefficients which yield the minimum error sums of squares SE where

$$SE = \hat{\dot{u}}_1'\hat{\dot{u}}_1$$

Calculation of the variance-covariance matrix of $\hat{\delta}_1$ is as

7. $\hat{\sigma}_{11} = SE_{min}/(T-K)$

$$Var(\hat{\delta}_1) = \sigma_{11}[(\hat{\dot{Y}}_1 \quad \dot{X}_1)'(\hat{\dot{Y}}_1 \quad \dot{X}_1)]^{-1}$$

Where

$$\hat{\dot{Y}}_1 = \dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y}_1$$

```
    ∇ GLIML
[ 1]  ABC2← 0 6 ρ0
[ 2]  BL2BAD←0ρ0
[ 3]  AVBC2← 0 5 ρ0
[ 4]  LL←RF←ER2←ι0
[ 5]  II←1
[ 6]  RT←(-1)+(ι19)÷10
[ 7] LL1:Y1←YDATA[(30×(II-1))+ι30;1]
[ 8]  Y2←YDATA[(30×(II-1))+ι30;2]
[ 9]  Y3←YDATA[(30×(II-1))+ι30;3]
[10]  I←1
[11]  ER2←ι0
[12] L1:R←RT[I]
[13]  X1R←(1 0 ↓X1)-R× ¯1 0 ↓X1
[14]  Y1R←(1↓Y1)-R×¯1↓Y1
[15]  Y2R←(1↓Y2)-R×¯1↓Y2
[16]  Y3R←(1↓Y3)-R×¯1↓Y3
[17]  YY←⍉(3 29 ρ(Y1R,Y2R,Y3R))
[18]  XT←(1 0 ↓Z)-R× ¯1 0 ↓Z
[19]  W1←((⍉YY)+.×YY)-((⍉YY)+.×X1R)+.×YY⌹X1R
[20]  W2←((⍉YY)+.×YY)-((⍉YY)+.×XT)+.×YY⌹XT
[21]  L←1+÷MXEV Q←(⌹(W1-W2))+.×W2
[22]  WW←W1-L×W2
[23]  BN←(-WW[;1])⌹ 0 1 ↓WW
[24]  B1←(-BN),(⌹(⍉X1R)+.×X1R)+.×(⍉X1R)+.×YY+.×1,BN
[25]  ER2←ER2,+/(ER←Y1R-(Y2R,Y3R,X1R)+.×B1)*2
[26]  →L1×ι(19≥I←I+1)
[27]  R1←(ER2∊⌊/ER2)/RT
[28] ⍝ FINAL ESTIMATION OF THE FIRST EQUATION
[29]  X1R←(1 0 ↓X1)-R1× ¯1 0 ↓X1
[30]  Y1R←(1↓Y1)-R1×¯1↓Y1
[31]  Y2R←(1↓Y2)-R1×¯1↓Y2
[32]  Y3R←(1↓Y3)-R1×¯1↓Y3
[33]  YY←⍉(3 29 ρ(Y1R,Y2R,Y3R))
[34]  XT←(1 0 ↓Z)-R1× ¯1 0 ↓Z
[35]  W1←((⍉YY)+.×YY)-((⍉YY)+.×X1R)+.×YY⌹X1R
[36]  W2←((⍉YY)+.×YY)-((⍉YY)+.×XT)+.×YY⌹XT
[37]  L←1+÷MXEV Q←(⌹(W1-W2))+.×W2
[38]  WW←W1-L×W2
[39]  BN←(-WW[;1])⌹ 0 1 ↓WW
[40]  B1←(-BN),(⌹(⍉X1R)+.×X1R)+.×(⍉X1R)+.×YY+.×1,BN
[41]  ABC2←ABC2,[1](B1,R1)
[42]  UH←Y1R-(Y2R,Y3R,X1R)+.×B1
[43]  S2←(+/UH*2)÷26
[44]  Y2HR←XT+.×Y2R⌹XT
[45]  Y3HR←XT+.×Y3R⌹XT
[46]  VAR←S2×⌹(⍉(Y2HR,Y3HR,X1R))+.×Y2HR,Y3HR,X1R
[47]  AVBC2←AVBC2,[1] VAR[1;1],VAR[2;2],VAR[3;3],VAR[4;4],VA
      R[5;5]
[48]  →LL1×ι(100≥II←II+1)
[49]  '***END'
    ∇
```

B. Systems with Lagged Endogenous Variable

1.Theil: Augmented Reduced Form

$Y B + Y_{-1}\Gamma_1 + X \Gamma_2 = U$

1. The first structural equation of the above system is

$$y_1 = Y_1 B_1 + y_{1,-1}\gamma_1 + X_1\gamma_2 + u_1 = Z_1\delta_1 + u_1$$

2. $W = (Y_{-1}\ Y_{-2}\ X\ X_{-1})$

Denote

$$\hat{W}_1 = (\hat{Y}_1\ \hat{y}_{1,-1}\ X_1)$$

$$\hat{Y}_1 = W(W'W)^{-1}W'Y_1$$

$$\hat{y}_1 = W(W'W)^{-1}W'y_1$$

$$Z_1 = (Y_1\ y_{1,-1}\ X_1)$$

$$\hat{\delta}_1 = (\hat{W}_1'Z_1)^{-1}\hat{W}_1'y_1$$

and $\hat{y}_{1,-1}$ is a one period lag of $\hat{y}_1$.

3. $\hat{u}_1 = y_1 - Z_1\hat{\delta}_1$

$$\hat{r} = \text{CORC formula}$$

4. Transform the first structural equation by the CORC method.

$$\bar{y}_1 = \bar{Y}_1 B_1 + \bar{y}_{1,-1}\gamma_1 + \bar{X}_1\gamma_2 + \epsilon_1$$

5. Obtain $\hat{\bar{Y}}_1$ by applying OLS to the transformed augmented reduced form

$$\hat{\bar{Y}}_1 = \bar{Z}_d\hat{\Pi}$$

Where

$$\bar{Z}_d = (\bar{Y}_{-1}\ \bar{Y}_{-2}\ \bar{X}\ \bar{X}_{-1})$$

$$\hat{\Pi} = (\bar{Z}_d'\bar{Z}_d)^{-1}\bar{Z}_d'\bar{Y}_1$$

6. Denote

$$\hat{\bar{Z}}_1 = (\hat{\bar{Y}}_1\ \bar{y}_{1,-1}\ \bar{X}_1)$$

$$\hat{\delta}_1 = (\hat{\overline{Z}}_1'\hat{\overline{Z}}_1)^{-1}\hat{\overline{Z}}_1'\overline{y}_1$$

7. Iterate steps 2, 3, 4, 5, and 6 until convergence.

8. Calculation of variance-covariance matrix is as before.

```
     ∇ THEILA
[ 1]   ATCA← 0 8 ρ0
[ 2]   AVTA← 0 6 ρ0
[ 3]   TABAD←0ρ0
[ 4]   II←1
[ 5]  L1:Y1←YDATA[(31×(II-1))+ι31;1]
[ 6]   Y2←YDATA[(31×(II-1))+ι31;2]
[ 7]   Y3←YDATA[(31×(II-1))+ι31;3]
[ 8]   Y11←1↓Y1
[ 9]   Y21←1↓Y2
[10]   Y31←1↓Y3
[11]   YL1←¯1↓Y1
[12]   YL2←¯1↓Y2
[13]   XI←YL1, 1 0 ↓X1
[14]   W←(¯1↓Y11),(¯1↓Y21),(¯1↓Y31),(¯1↓YL1),(¯1↓YL2),(2 0 ↓Z
       ), 1 0 ↓ ¯1 0 ↓ 0 ¯1 ↓Z
[15]   Y1H←W+.×(1↓Y11)⌹W
[16]   Y2H←W+.×(1↓Y21)⌹W
[17]   Y3H←W+.×(1↓Y31)⌹W
[18]   WH←(1↓Y2H),(1↓Y3H),(¯1↓Y1H), 3 0 ↓X1
[19]   M←(2↓Y21),(2↓Y31),(2↓YL1), 3 0 ↓X1
[20]   B←(⌹((⍉WH)+.×M))+.×(⍉WH)+.×2↓Y11
[21]   R←I←0
[22]  L2:R1←R
[23]   R←(1↓E)⌹¯1↓E←Y11-(Y21,Y31,XI)+.×B
[24]   TABAD←TABAD,II×ι(R≥1)
[25]   R←(R,0.99)[1+R≥1]
[26]   Y1R←(1↓Y11)-R×¯1↓Y11
[27]   Y2R←(1↓Y21)-R×¯1↓Y21
[28]   Y3R←(1↓Y31)-R×¯1↓Y31
[29]   XIR←(1 0 ↓XI)-R× ¯1 0 ↓XI
[30]   WR←(1 0 ↓W)-R× ¯1 0 ↓W
[31]   Y2HR←WR+.×(1↓Y2R)⌹WR
[32]   Y3HR←WR+.×(1↓Y3R)⌹WR
[33]   B←(1↓Y1R)⌹Y2HR,Y3HR, 1 0 ↓XIR
[34]   →L2×ι(20≥I←I+1)∧(0.005≤|R-R1)
[35]   ATCA←ATCA,[1](B,R,I)
[36]   URH←Y1R-(Y2R,Y3R,XIR)+.×B
[37]   S2←(+/URH*2)÷24
[38]   V←S2×⌹(⍉(Y2HR,Y3HR, 1 0 ↓XIR))+.×Y2HR,Y3HR, 1 0 ↓XIR
[39]   AVTA←AVTA,[1] V[1;1],V[2;2],V[3;3],V[4;4],V[5;5],V[6;6
       ]
[40]   →L1×ι(100≥II←II+1)
[41]   '***END'
     ∇
```

2. **Theil: Ordinary Reduced Form**

$$Y B + Y_{-1}\Gamma_1 + X \Gamma_2 = U$$

The first structural equation is

$$y_1 = Y_1 B_1 + y_{1,-1}\gamma_1 + X_1\gamma_2 + u_1 = Z_1\delta_1 + u_1$$

1. $W = (Y_{-1} \; Y_{-2} \; X \; X_{-1})$

Denote

$$\hat{W}_1 = (\hat{Y}_1 \; \hat{y}_{1,-1} \; X_1)$$

$$Z_1 = (Y_1 \; y_{1,-1} \; X_1)$$

$$\hat{\delta}_1 = (\hat{W}_1'Z_1)^{-1}\hat{W}_1'y_1$$

Where

$$\hat{Y}_1 = W(W'W)^{-1}W'Y_1$$

$$\hat{y}_1 = W(W'W)^{-1}W'y_1$$

2. $\hat{u}_1 = y_1 - Z_1\hat{\delta}_1$

$\hat{r}$ = Prais-Winsten formula

3. Transform the first structural equation by Prais-Winsten, P, matrix

$$Py_1 = PY_1B_1 + Py_{1,-1}\gamma_1 + PX_1\gamma_2 + \epsilon_1$$

or

$$\dot{y}_1 = \dot{Y}_1B_1 + \dot{y}_{1,-1}\gamma_1 + \dot{X}_1\gamma_2 + \epsilon_1$$

4. Write the transformed ordinary reduced form as

$$P\,Y = P\,Y_{-1}\Pi_0 + P\,X\,\Pi_1 + \dot{V}$$

To obtain fitted values of PY, we have to estimate the reduced form by instrumental variable estimator, using IV for Y_{-1} such as

$$\hat{Y}_{-1} = X_{-1}\hat{\Pi}$$

Where

$$\hat{\Pi} = (X_{-1}'X_{-1})^{-1}X_{-1}'Y_{-1}$$

Denote

$$\dot{W}_d = (P\ \hat{Y}_{-1}\ \ P\ X)$$

$$\dot{Z} = (P\ Y_{-1}\ \ P\ X)$$

$$\hat{\Pi}_{IV} = (\dot{W}_d'\dot{Z})^{-1}\dot{W}_d'\dot{Y}$$

and obtain

$$\widehat{PY}_1 = \hat{\dot{Y}}_1 = \dot{Z}\ \hat{\Pi}_{IV}$$

Denote

$$\hat{\dot{Z}}_1 = (\hat{\dot{Y}}_1\ \dot{y}_{1,-1}\ \dot{X}_1)$$

$$\hat{\delta}_1 = (\hat{\dot{Z}}_1'\hat{\dot{Z}}_1)^{-1}\hat{\dot{Z}}_1'\dot{y}_1$$

6. Iterate steps 2, 3, 4, and 5 until convergence.

7. Calculation of the variance-covariance matrix

$$\hat{\dot{u}}_1 = \dot{y}_1 - \dot{Z}_1\hat{\delta}_1$$

$$\hat{\sigma}_{11} = \Sigma\hat{\dot{u}}_1'\hat{\dot{u}}_1/(T-K)$$

$$Var(\hat{\delta}_1) = \hat{\sigma}_{11}(\hat{\dot{Z}}_1'\hat{\dot{Z}}_1)^{-1}$$

```
    ∇ THEILO
  TC1← 0 8 ρ0
  VTC1← 0 6 ρ0
  TBAD←0ρ0
  II←1
L1:Y1←YDATA[(31×(II-1))+⍳31;1]
  Y2←YDATA[(31×(II-1))+⍳31;2]
  Y3←YDATA[(31×(II-1))+⍳31;3]
  Y11←1↓Y1
  Y21←1↓Y2
  Y31←1↓Y3
  YL1←¯1↓Y1
  YL2←¯1↓Y2
  XI←YL1, 1 0 ↓X1
  YL1H←(¯1 0 ↓Z)+.×YL1⌹(¯1 0 ↓Z)
  YL2H←(¯1 0 ↓Z)+.×YL2⌹(¯1 0 ↓Z)
  W←(¯1↓Y11),(¯1↓Y21),(¯1↓Y31),(¯1↓YL1),(¯1↓YL2),(2 0 ↓Z
  ), 1 0 ↓ ¯1 0 ↓ 0 ¯1 ↓Z
  Y1H←W+.×(1↓Y11)⌹W
  Y2H←W+.×(1↓Y21)⌹W
  Y3H←W+.×(1↓Y31)⌹W
  WH←(1↓Y2H),(1↓Y3H),(¯1↓Y1H), 3 0 ↓X1
  M←(2↓Y21),(2↓Y31),(2↓YL1), 3 0 ↓X1
  Z1← 1 0 ↓Z
  B←(⌹((⍉WH)+.×M))+.×(⍉WH)+.×2↓Y11
  R←I←0
L2:R1←R
  R←((1↓E)+.×¯1↓E)÷+/(1↓¯1↓E←Y11-(Y21,Y31,XI)+.×B)*2
  TBAD←TBAD,II×⍳(R≥1)
  R←(R,0.99)[1+R≥1]
  RR←(1-R*2)*0.5
⍝   ***TRANSFORMATION OF THE VARIABLES
  Y1R←(Y11[1]×RR),(1↓Y11)-R×¯1↓Y11
  Y2R←(Y21[1]×RR),(1↓Y21)-R×¯1↓Y21
  Y3R←(Y31[1]×RR),(1↓Y31)-R×¯1↓Y31
  ZR← 1 7 ρZ1[1;]×RR
  ZR←ZR,[1](1 0 ↓Z1)-R× ¯1 0 ↓Z1
  XIR← 1 4 ρXI[1;]×RR
  XIR←XIR,[1](1 0 ↓XI)-R× ¯1 0 ↓XI
  YL1R←(YL1[1]×RR),(1↓YL1)-R×¯1↓YL1
  YL2R←(YL2[1]×RR),(1↓YL2)-R×¯1↓YL2
  YL1HR←(YL1H[1]×RR),(1↓YL1H)-R×¯1↓YL1H
  YL2HR←(YL2H[1]×RR),(1↓YL2H)-R×¯1↓YL2H
  WR←(YL1HR,YL2HR,ZR)
  MR←(YL1R,YL2R,ZR)
  Y2RH←MR+.×(⌹((⍉WR)+.×MR))+.×(⍉WR)+.×Y2R
  Y3RH←MR+.×(⌹((⍉WR)+.×MR))+.×(⍉WR)+.×Y3R
  B←Y1R⌹(Y2RH,Y3RH,XIR)
  →L2×⍳(20≥I←I+1)∧(0.005≤|R-R1)
  TC1←TC1,[1](B,R,I)
  URH←Y1R-(Y2R,Y3R,XIR)+.×B
  S2←(+/URH*2)÷24
  VA←S2×⌹(⍉(Y2RH,Y3RH,XIR))+.×Y2RH,Y3RH,XIR
  VTC1←VTC1,[1] VA[1;1],VA[2;2],VA[3;3],VA[4;4],VA[5;5],
```

```
       VA[6;6]
[53]   →L1×ι(100≥II←II+1)
[54]   '***END'
     ∇
```

3. **Fair Estimator**

$y_1 = Y_1B_1 + X_1*\gamma_1 + u_1 = Z_1\delta_1 + u_1$

Where

$X_1* = (y_{1,-1} \quad X_1)$

1. $W = (X_1* \; X_{1,-1}* \; y_{1,-1} \quad Y_{1,-1})$

Note that some of the lagged endogenous variables are included in $X_{1,-1}*$, $y_{1,-1}$ and $Y_{1,-1}$. However, it is clear that in the matrix W we will count them only once. Denote

$\hat{W}_1 = (\hat{Y}_1 \; \hat{y}_{1,-1} \; X_1)$

$Z_1 = (Y_1 \; y_{1,-1} \; X_1)$

$\hat{\delta}_1 = (\hat{W}_1'Z_1)^{-1}\hat{W}_1'y_1$

$\hat{Y}_1 = W(W'W)^{-1}W'Y_1$

$\hat{y}_1 = W(W'W)^{-1}W'y_1$

2. $\hat{u}_1 = y_1 - Z_1\hat{\delta}_1$

$\hat{r}$ = CORC formula

3. Transforming the structural equation

$\bar{y}_1 = y_1 - \hat{r}y_{1,-1}$

$\bar{X}_1* = X_1* - \hat{r}X_{1,-1}*$

$\hat{\bar{Y}}_1 = Y_1 - \hat{r}Y_{1,-1}$

Denote

$\hat{\bar{Z}}_1 = (\hat{\bar{Y}}_1 \quad \bar{X}_1*)$

4. $\hat{\delta}_1 = (\hat{\bar{Z}}_1'\hat{\bar{Z}}_1)^{-1}\hat{\bar{Z}}_1'\bar{y}_1$

5. Iterate steps 2, 3, and 4 until convergence.

Calculation of the variance-covariance matrix

$\hat{\bar{u}}_1 = \bar{y}_1 - \bar{Z}_1\hat{\delta}_1$

$\hat{\sigma}_{11} = \Sigma\hat{\bar{u}}_1'\hat{\bar{u}}_1/(T-K)$

$$\mathrm{Var}(\hat{\delta}_1) = \hat{\sigma}_{11}(\hat{\bar{Z}}_1{}'\hat{\bar{Z}}_1)^{-1}$$

```
     ∇ FAIR
[ 1]   AFC← 0 8 ρ0
[ 2]   AVFC← 0 6 ρ0
[ 3]   FBAD←0ρ0
[ 4]   II←1
[ 5]  L1:Y1←YDATA[(31×(II-1))+ι31;1]
[ 6]   Y2←YDATA[(31×(II-1))+ι31;2]
[ 7]   Y3←YDATA[(31×(II-1))+ι31;3]
[ 8]   Y11←1↓Y1
[ 9]   Y21←1↓Y2
[10]   Y31←1↓Y3
[11]   YL←¯1↓Y1
[12]   XI←YL, 1 0 ↓X1
[13]   IV← ¯1 0 ↓Y11,Y21,Y31,XI
[14]   V←IV, 2 0 ↓X11,X44
[15]   Y1H←V+.×(1↓Y11)⌹V
[16]   Y2H←V+.×(1↓Y21)⌹V
[17]   Y3H←V+.×(1↓Y31)⌹V
[18]   W←(1↓Y2H),(1↓Y3H),(¯1↓Y1H), 3 0 ↓X1
[19]   M←(2↓Y21),(2↓Y31),(2↓YL), 3 0 ↓X1
[20]   B←(⌹((⍉W)+.×M))+.×(⍉W)+.×2↓Y11
[21]   R←I←0
[22]  L2:R1←R
[23]   R←(1↓E)⌹¯1↓E←Y11-(Y21,Y31,XI)+.×B
[24]   FBAD←FBAD,II×ι(R≥1)
[25]   R←(R,0.99)[1+R≥1]
[26]   XIR←(1 0 ↓XI)-R× ¯1 0 ↓XI
[27]   Y1R←(1↓Y11)-R×¯1↓Y11
[28]   Y2R←(1↓Y21)-R×¯1↓Y21
[29]   Y3R←(1↓Y31)-R×¯1↓Y31
[30]   Y2HR←Y2H-R×¯1↓Y21
[31]   Y3HR←Y3H-R×¯1↓Y31
[32]   B←Y1R⌹(Y2HR,Y3HR,XIR)
[33]   →L2×ι(20≥I←I+1)∧(0.005≤|R-R1)
[34]   AFC←AFC,[1](B,R,I)
[35]   URH←Y1R-(Y2R,Y3R,XIR)+.×B
[36]   S2←(+/URH*2)÷24
[37]   VA←S2×⌹(⍉(Y2HR,Y3HR,XIR))+.×Y2HR,Y3HR,XIR
[38]   AVFC←AVFC,[1] VA[1;1],VA[2;2],VA[3;3],VA[4;4],VA[5;5],
       VA[6;6]
[39]   →L1×ι(100≥II←II+1)
[40]   '***END'
     ∇
```

4. **Dhrymes Estimator (C2SLSA)**

$y_1 = Y_1B_1 + y_{1,-1}\gamma_1 + X_1\gamma_2 + u_1 = Z_1\delta_1 + u_1$

1. Estimate the structural equation under consideration by an instrumental variable estimator, as in the Theil(ordinary reduced form), and calculate a consistent estimate of the autocorrelation coefficient r.

2. Obtain fitted values of Y_1 by application of OLS to the augmented reduced form

$$\hat{Y}_1 = Y_{-1}\hat{\Pi}_1 + Y_{-2}\hat{\Pi}_2 \ X \ \hat{\Pi}_3 + X_{-1}\hat{\Pi}_4$$

3. Transform the first equation as

$$y_1-\hat{r}y_{1,-1} = [(Y_1-\hat{r}Y_{1,-1}) \quad (y_{1,-1}-\hat{r}y_{1,-2}) \ (X_1-\hat{r}X_{1,-1})$$
$$]\delta_1 + \epsilon_1$$

4. Denote

$$\hat{\bar{Z}}_1 = [(\hat{Y}_1-\hat{r}Y_{1,-1}) \ (y_{1,-1}-\hat{r}y_{1,-2}) \ (X_1-\hat{r}X_{1,-1})]$$

$$\hat{\delta}_1 = (\hat{\bar{Z}}_1'\hat{\bar{Z}}_1)^{-1}\hat{\bar{Z}}_1'\bar{y}_1$$

5. Use $\hat{\delta}_1$ to obtain new estimate of r and iterate steps 3 and 4 until convergence.

Calculation of the variance-covariance matrix

$$\hat{\bar{u}}_1 = \bar{y}_1 - \bar{Z}_1\hat{\delta}_1$$

$$\hat{\sigma}_{11} = \Sigma\hat{\bar{u}}_1'\hat{\bar{u}}_1/(T-K)$$

$$Var(\hat{\delta}_1) = \hat{\sigma}_{11}(\hat{\bar{Z}}_1'\hat{\bar{Z}}_1)^{-1}$$

```
    ∇ DHRYMIT
  AD2S← 0 8 ρ0
  AVD2S← 0 6 ρ0
  DB2S←0ρ0
  II←1
L1:Y1←YDATA[(31×(II-1))+ι31;1]
  Y2←YDATA[(31×(II-1))+ι31;2]
  Y3←YDATA[(31×(II-1))+ι31;3]
  Y11←1↓Y1
  Y21←1↓Y2
  Y31←1↓Y3
  YL1←¯1↓Y1
  YL2←¯1↓Y2
  XI←YL1, 1 0 ↓X1
  W←(¯1↓Y11),(¯1↓Y21),(¯1↓Y31),(¯1↓YL1),(¯1↓YL2),(2 0 ↓Z
  ), 1 0 ↓ ¯1 0 ↓ 0 ¯1 ↓Z
  Y1H←W+.×(1↓Y11)⌹W
  Y2H←W+.×(1↓Y21)⌹W
  Y3H←W+.×(1↓Y31)⌹W
  WW←(1↓Y2H),(1↓Y3H),(¯1↓Y1H), 3 0 ↓X1
  ZR←(2↓Y21),(2↓Y31),(2↓YL1), 3 0 ↓X1
  B←(⌹((⍉WW)+.×ZR))+.×(⍉WW)+.×2↓Y11
  R←I←0
L2:R1←R
  R←(1↓E)⌹¯1↓E←Y11-(Y21,Y31,XI)+.×B
  DB2S←DB2S,II×ι(R≥1)
  R←(R,0.99)[1+R≥1]
⍝ ****TRANSFORMATION OF THE VARIABLES
  XIR←(1 0 ↓XI)-R× ¯1 0 ↓XI
  Y1R←(1↓Y11)-R×¯1↓Y11
  Y2R←(1↓Y21)-R×¯1↓Y21
  Y3R←(1↓Y31)-R×¯1↓Y31
  Y2HR←Y2H-R×¯1↓Y21
  Y3HR←Y3H-R×¯1↓Y31
  B←Y1R⌹(Y2HR,Y3HR,XIR)
  →L2×ι(20≥I←I+1)∧(0.005≤|R-R1)
  AD2S←AD2S,[1](B,R,I)
  URH←Y1R-(Y2R,Y3R,XIR)+.×B
  S2←(+/URH*2)÷24
  V←S2×⌹(⍉(Y2HR,Y3HR,XIR))+.×Y2HR,Y3HR,XIR
  AVD2S←AVD2S,[1] V[1;1],V[2;2],V[3;3],V[4;4],V[5;5],V[6
  ;6]
  →L1×ι(100≥II←II+1)
  '***END'
    ∇
```

5. **Dhrymes Two Step Instrumental Variable Estimator**

$Y\,B + Y_{-1}\Gamma_1 + X\,\Gamma_2 = U$

The first equation can be written as

$$y_1 = Y_1 B_1 + y_{1,-1}\gamma_1 + X_1\gamma_2 + u_1 = Z_1\delta_1 + u_1$$

The ordinary reduced form of the above system is

$$Y_t = Y_{-1}\Gamma_1 B^{-1} + X\,\Gamma_2 B^{-1} + UB^{-1} = Y_{-1}\Pi_1 + X\,\Pi_2 + V$$

1. Estimate all structural equations by IV and use the estimated coefficients to form a consistent estimate of Π_1 and Π_2. Then obtain the fitted values of Y_1 using $Y_{-1}(0)=0$ as the initial condition.

$$\hat{Y}_1 = Y_{-1}\hat{\Pi}_1 + X\,\hat{\Pi}_2$$

2. Use the initial IV estimates of the structural equation under consideration to calculate $\hat{r}$.

$$\hat{u}_1 = y_1 - Z_1\hat{\delta}_1$$

$\hat{r}$ = Prais-Winsten formula

3. Define

$$\hat{W}_1 = (P\hat{Y}_1\ \ Py_{1,-1}\ \ PX_1)$$

$$\dot{Z}_1 = (PY_1\ \ Py_{1,-1}\ \ PX_1)$$

and estimate

$$\hat{\delta}_1 = (\hat{W}_1{}'\mathring{Z}_1)^{-1}\hat{W}_1{}'\mathring{y}_1$$

Calculation of variance-covariance matrix

4. $\hat{\mathring{u}}_1 = \mathring{y}_1 - \mathring{Z}_1\hat{\delta}_1$

$$\hat{\sigma}_{11} = \Sigma\hat{\mathring{u}}_1{}'\hat{\mathring{u}}_1/(T-K)$$

$$\mathrm{Var}(\hat{\delta}_1) = \hat{\sigma}_{11}(\hat{W}_1{}'\dot{Z}_1)^{-1}$$

```
    ∇ DHRYM2S
[ 1]  DCI← 0 8 ρ0
[ 2]  DC2← 0 7 ρ0
[ 3]  VDC2← 0 6 ρ0
[ 4]  VDCI← 0 6 ρ0
[ 5]  YH1←YH2←YH3←32ρ0
[ 6]  DBAD←0ρ0
[ 7]  II←1
[ 8] L1:Y1←YDATA[(31×(II-1))+ι31;1]
[ 9]  Y2←YDATA[(31×(II-1))+ι31;2]
[10]  Y3←YDATA[(31×(II-1))+ι31;3]
[11]  Y11←1↓Y1
[12]  Y21←1↓Y2
[13]  Y31←1↓Y3
[14]  YL1←¯1↓Y1
[15]  YL2←¯1↓Y2
[16]  XI←YL1, 1 0 ↓X1
[17]  XI1← 1 0 ↓X1
[18]  W←(¯1↓Y11),(¯1↓Y21),(¯1↓Y31),(¯1↓YL1),(¯1↓YL2),(2 0 ↓Z
      ), 1 0 ↓ ¯1 0 ↓ 0 ¯1 ↓Z
[19]  Y1H←W+.×(1↓Y11)⌹W
[20]  Y2H←W+.×(1↓Y21)⌹W
[21]  Y3H←W+.×(1↓Y31)⌹W
[22]  W1←(1↓Y2H),(1↓Y3H),(¯1↓Y1H), 3 0 ↓X1
[23]  W2←(1↓Y1H),(¯1↓Y2H), 3 0 ↓X2
[24]  Z1←(2↓Y21),(2↓Y31),(2↓YL1), 3 0 ↓X1
[25]  Z2←(2↓Y11),(2↓YL2), 3 0 ↓X2
[26]  B1←(⌹((⍉W1)+.×Z1))+.×(⍉W1)+.×2↓Y11
[27]  B2←(⌹((⍉W2)+.×Z2))+.×(⍉W2)+.×2↓Y21
[28]  B3←(1↓Y31)⌹(Y2H, 2 0 ↓X3)
[29]  BB← 3 3 ρ1,(-B1[1]),(-B1[2]),(-B2[1]),1,0,0,(-B3[1]),1
[30]  Q1← 1 7 ρB1[4],0,0,B1[5],0,0,B1[6]
[31]  Q2← 1 7 ρ0,B2[3],0,B2[4],0,B2[5],B2[6]
[32]  Q3← 1 7 ρ0,B3[2],B3[3],0,B3[4],0,B3[5]
[33]  Q←Q1,[1] Q2,[1] Q3
[34]  P← 3 3 ρB1[3],0,0,0,B2[2],0,0,0,0
[35]  PP1←(⌹BB)+.×P
[36]  PP2←(⌹BB)+.×Q
[37]  I←2
[38]  YH1←YH2←YH3←32ρ0
[39] L2:YH1[I]←PP1[1;]+.×(YH1[I-1],YH2[I-1],YH3[I-1])+PP2[1;
      ]+.×⍉Z[I-1;]
[40]  YH2[I]←PP1[2;]+.×(YH1[I-1],YH2[I-1],YH3[I-1])+PP2[2;]+
      .×⍉Z[I-1;]
[41]  YH3[I]←PP1[3;]+.×(YH1[I-1],YH2[I-1],YH3[I-1])+PP2[3;]+
      .×⍉Z[I-1;]
[42]  →L2×ι(32≥I←I+1)
[43]  R←((1↓E)+.×¯1↓E)÷+/(1↓¯1↓E←Y11-(Y21,Y31,XI)+.×B1)*2
[44]  RR←(1-R*2)*0.5
[45]  YHL←1↓¯1↓YH1
[46]  YH1←2↓YH1
[47]  YH2←2↓YH2
[48]  YH3←2↓YH3
[49]  YHLR←(YHL[1]×RR),(1↓YHL)-R×¯1↓YHL
```

```
[50]  YH2R←(YH2[1]×RR),(1↓YH2)-R×¯1↓YH2
[51]  YH3R←(YH3[1]×RR),(1↓YH3)-R×¯1↓YH3
[52]  Y1R←(Y11[1]×RR),(1↓Y11)-R×¯1↓Y11
[53]  Y2R←(Y21[1]×RR),(1↓Y21)-R×¯1↓Y21
[54]  Y3R←(Y31[1]×RR),(1↓Y31)-R×¯1↓Y31
[55]  X1R← 1 3 ρXI1[1;]×RR
[56]  X1R←X1R,[1](1 0 ↓XI1)-R× ¯1 0 ↓XI1
[57]  XIR← 1 4 ρXI[1;]×RR
[58]  XIR←XIR,[1](1 0 ↓XI)-R× ¯1 0 ↓XI
[59]  YL1R←(YL1[1]×RR),(1↓YL1)-R×¯1↓YL1
[60]  WW←YH2R,YH3R,YHLR,X1R
[61]  ZR←Y2R,Y3R,YL1R,X1R
[62]  B←(⌹((⍉WW)+.×ZR))+.×(⍉WW)+.×Y11
[63]  DC2←DC2,[1] B,R
[64]  URH←Y1R-(Y2R,Y3R,XIR)+.×B
[65]  S2←(+/URH*2)÷24
[66]  VA←S2×⌹(⍉(YH2R,YH3R,YHLR,X1R))+.×Y2R,Y3R,YL1R,X1R
[67]  VDC2←VDC2,[1] VA[1;1],VA[2;2],VA[3;3],VA[4;4],VA[5;5],
      VA[6;6]
[68]  I←1
[69] L3:R1←R
[70]  R←((1↓E)+.×¯1↓E)÷+/(1↓¯1↓E←Y11-(Y21,Y31,XI)+.×B)*2
[71]  DBAD←DBAD,II×ι(R≥1)
[72]  R←(R,0.99)[1+R≥1]
[73]  RR←(1-R*2)*0.5
[74]  YHLR←(YHL[1]×RR),(1↓YHL)-R×¯1↓YHL
[75]  YH2R←(YH2[1]×RR),(1↓YH2)-R×¯1↓YH2
[76]  YH3R←(YH3[1]×RR),(1↓YH3)-R×¯1↓YH3
[77]  Y1R←(Y11[1]×RR),(1↓Y11)-R×¯1↓Y11
[78]  Y2R←(Y21[1]×RR),(1↓Y21)-R×¯1↓Y21
[79]  Y3R←(Y31[1]×RR),(1↓Y31)-R×¯1↓Y31
[80]  X1R← 1 3 ρXI1[1;]×RR
[81]  X1R←X1R,[1](1 0 ↓XI1)-R× ¯1 0 ↓XI1
[82]  XIR← 1 4 ρXI[1;]×RR
[83]  XIR←XIR,[1](1 0 ↓XI)-R× ¯1 0 ↓XI
[84]  YL1R←(YL1[1]×RR),(1↓YL1)-R×¯1↓YL1
[85]  WW←YH2R,YH3R,YHLR,X1R
[86]  ZR←Y2R,Y3R,YL1R,X1R
[87]  B←(⌹((⍉WW)+.×ZR))+.×(⍉WW)+.×Y11
[88]  →L3×ι(20≥I←I+1)∧(0.005≤|R-R1)
[89]  DCI←DCI,[1] B,R,I
[90]  URH←Y1R-(Y2R,Y3R,XIR)+.×B
[91]  S2←(+/URH*2)÷24
[92]  V←S2×⌹(⍉(YH2R,YH3R,YHLR,X1R))+.×Y2R,Y3R,YL1R,X1R
[93]  VDCI←VDCI,[1] V[1;1],V[2;2],V[3;3],V[4;4],V[5;5],V[6;6
      ]
[94]  →L1×ι(100≥II←II+1)
[95]  '***END'
    ∇
```

6. **Hatanaka(A) Estimator**

$y_1 = Y_1B_1 + y_{1,-1}\gamma_1 + X_1\gamma_2 + u_1 = Z_1\delta_1 + u_1$

1. Applying OLS to the augmented reduced form and obtain a fitted values of the Y_1 and y_1

Denote

$$W_1 = (\hat{Y}_1 \; \hat{y}_{1,-1} \; X_1)$$

$$Z_1 = (Y_1 \; y_{1,-1} \; X_1)$$

2. Apply IV to the structural equation under consideration and obtain fitted values of $\hat{u}_1$ and thus the $\hat{r}$

$$\hat{\delta}_{IV} = (W_1'Z_1)^{-1}W_1'y_1$$

3. $\hat{u}_1 = y_1 - Z_1\hat{\delta}_{IV}$

$$\hat{r} = \text{CORC formula}$$

4. Denote

$$\hat{\bar{Z}}_1 = [(\hat{Y}_1-\hat{r}Y_{1,-1}) \; (y_{1,-1}-\hat{r}y_{1,-2}) \; (X_1-\hat{r}X_{1,-1}) \; \hat{u}_{1,-1}]$$

$$\bar{y}_1 = y_1 - \hat{r}y_{1,-1}$$

5. $\begin{pmatrix} \hat{\delta}_1 \\ \hat{r}_{11} \end{pmatrix} = (\hat{\bar{Z}}_1'\hat{\bar{Z}}_1)^{-1}\hat{\bar{Z}}_1'\bar{y}_1$

and finally

$$\begin{pmatrix} \hat{\delta}_1 \\ \hat{r}_1 \end{pmatrix} = \begin{pmatrix} \hat{\delta}_1 \\ \hat{r}+\hat{r}_{11} \end{pmatrix}$$

Calculation of the variance-covariance

6. $\hat{\bar{u}}_1 = \bar{y}_1 - \bar{Z}_1\hat{\delta}_1$

$$\hat{\sigma}_{11} = \Sigma\hat{\bar{u}}_1'\hat{\bar{u}}_1/(T-K)$$

$$\text{Var}\begin{pmatrix} \hat{\delta}_1 \\ \hat{r}_1 \end{pmatrix} = \hat{\sigma}_{11}(\hat{\bar{Z}}_1'\hat{\bar{Z}}_1)^{-1}$$

```
    ∇ HATANAKAA
[ 1]   AHA← 0 7 ρ0
[ 2]   AVHA← 0 7 ρ0
[ 3]   II←1
[ 4]  L1:Y1←YDATA[(31×(II-1))+⍳31;1]
[ 5]   Y2←YDATA[(31×(II-1))+⍳31;2]
[ 6]   Y3←YDATA[(31×(II-1))+⍳31;3]
[ 7]   Y11←1↓Y1
[ 8]   Y21←1↓Y2
[ 9]   Y31←1↓Y3
[10]   YL1←¯1↓Y1
[11]   YL2←¯1↓Y2
[12]   XI←YL1, 1 0 ↓X1
[13]   W←(¯1↓Y11),(¯1↓Y21),(¯1↓Y31),(¯1↓YL1),(¯1↓YL2),(2 0 ↓Z
       ), 1 0 ↓ ¯1 0 ↓ 0 ¯1 ↓Z
[14]   Y1H←W+.×(1↓Y11)⌹W
[15]   Y2H←W+.×(1↓Y21)⌹W
[16]   Y3H←W+.×(1↓Y31)⌹W
[17]   W1←(1↓Y2H),(1↓Y3H),(¯1↓Y1H), 3 0 ↓X1
[18]   M←(2↓Y21),(2↓Y31),(2↓YL1), 3 0 ↓X1
[19]   B←(⌹((⍉W1)+.×M))+.×(⍉W1)+.×2↓Y11
[20]   UH←Y11-(Y21,Y31,XI)+.×B
[21]   R←(1↓UH)⌹¯1↓UH
[22]   Y2HR←Y2H-R×¯1↓Y21
[23]   Y3HR←Y3H-R×¯1↓Y31
[24]   XIR←(1 0 ↓XI)-R× ¯1 0 ↓XI
[25]   Y1R←(1↓Y11)-R×¯1↓Y11
[26]   Y2R←(1↓Y21)-R×¯1↓Y21
[27]   Y3R←(1↓Y31)-R×¯1↓Y31
[28]   WH←Y2HR,Y3HR,XIR,¯1↓UH
[29]   BR←Y1R⌹WH
[30]   AHA←AHA,[1] BR+(0,0,0,0,0,0,R)
[31]   URH←Y1R-(Y2R,Y3R,XIR)+.×¯1↓BR
[32]   S2←(+/URH*2)÷24
[33]   V←S2×⌹(⍉WH)+.×WH
[34]   AVHA←AVHA,[1] V[1;1],V[2;2],V[3;3],V[4;4],V[5;5],V[6;6
       ],V[7;7]
[35]   →L1×⍳(100≥II←II+1)
[36]   '***END'
    ∇
```

7. **Hatanaka(B) Estimator**

$y_1 = Y_1 B_1 + y_{1,-1}\gamma_1 + X_1\gamma_2 + u_1 = Z_1\delta_1 + u_1$

The augmented reduced form of the system is

$$Y = Y_{-1}\Pi_1 + Y_{-2}\Pi_2 + X\,\Pi_3 + X_{-1}\Pi_4 + V$$

1. Estimate all the structural equations by IV, and use the estimated coefficients to form a consistent estimate of Π_1, Π_2, Π_3, and Π_4. Then calculate the fitted values of Y as

$$\hat{Y} = Y_{-1}\hat{\Pi}_1 + Y_{-2}\hat{\Pi}_2 + X\,\hat{\Pi}_3 + X_{-1}\hat{\Pi}_4$$

Where $\hat{\Pi}_i$'s are formed using the consistent estimate of the structural coefficients obtained in the first step.

2. From the first stage IV estimation, calculate the fitted values of u_1. Then obtain a estimate of the autocorrelation coefficient r.

$$\hat{u}_1 = y_1 - Z_1\hat{\delta}_{IV}$$

$\hat{r}$ = CORC formula

3. Denote

$$\hat{\bar{W}}_1 = [(\hat{Y}_1 - \hat{r}Y_{1,-1})\ \ (y_{1,-1}-\hat{r}y_{1,-2})\ \ (X_1-\hat{r}X_{1,-1})\ \ \hat{u}_{1,-1}]$$

$$\bar{Z}_1 = [(Y_1-\hat{r}Y_{1,-1})\ \ (y_{1,-1}-\hat{r}y_{1,-2})\ \ (X_1-\hat{r}X_{1,-1})\ \ \hat{u}_{1,-1}]$$

Then estimate the structural coefficients as

$$\begin{pmatrix} \hat{\delta}_1 \\ \\ \hat{r}_{11} \end{pmatrix} = (\hat{\bar{W}}_1{}'\bar{Z}_1)^{-1}\hat{\bar{W}}_1{}'\bar{y}_1$$

Finally, the proposed estimate are

$$\begin{pmatrix} \hat{\delta}_1 \\ \\ \hat{r}_1 \end{pmatrix} = \begin{pmatrix} \hat{\delta}_1 \\ \\ \hat{r}_{11} + \hat{r} \end{pmatrix}$$

The variance-covariance matrix is

$$\mathrm{Var}\begin{pmatrix} \hat{\delta}_1 \\ \hat{r}_1 \end{pmatrix} = \hat{\sigma}_{11}(\hat{\bar{W}}_1{}'\bar{Z}_1)^{-1}$$

8. **Hatanaka(C) Estimator**

$y_1 = Y_1 B_1 + Y_{-1}\gamma_1 + X_1\gamma_2 + u_1$

1. Obtain W_1 as in Hatanaka(B).

2. Also, from the first stage IV estimation obtain

$$\hat{\hat{u}}_1 = \hat{u}_1 - \hat{r}\hat{u}_{1,-1}$$

3. Estimate the structural coefficient as

$$\begin{pmatrix} \tilde{\delta}_1 \\ \\ \hat{r}_{11} \end{pmatrix} = (\hat{\bar{W}}_1{}'\hat{\bar{W}}_1)^{-1}\hat{\bar{W}}_1{}'\hat{\bar{u}}_1$$

Then the final estimated coefficient are

$$\begin{pmatrix} \hat{\delta}_1 \\ \\ \hat{r}_{11} \end{pmatrix} = \begin{pmatrix} \tilde{\delta}_1 + \hat{\delta}_{IV} \\ \\ \hat{r} + \hat{r}_{11} \end{pmatrix}$$

Where $\hat{\delta}_{IV}$ is obtained from the first stage IV estimation.

4. The variance-covariance matrix of the structural

coefficient is

$$\mathrm{Var}\begin{pmatrix} \hat{\delta}_1 \\ \hat{r}_1 \end{pmatrix} = \hat{\sigma}_{11}(\hat{\bar{W}}_1{}'\hat{\bar{W}}_1)^{-1}$$

```
     ∇ HATANAKABC
[ 1]   AHB← 0 7 ⍴0
[ 2]   AHC← 0 7 ⍴0
[ 3]   AVHB← 0 7 ⍴0
[ 4]   AVHC← 0 7 ⍴0
[ 5]   II←1
[ 6]  L1:Y1←YDATA[(31×(II-1))+⍳31;1]
[ 7]   Y2←YDATA[(31×(II-1))+⍳31;2]
[ 8]   Y3←YDATA[(31×(II-1))+⍳31;3]
[ 9]   Y11←1↓Y1
[10]   Y21←1↓Y2
[11]   Y31←1↓Y3
[12]   YL1←¯1↓Y1
[13]   YL2←¯1↓Y2
[14]   XI←YL1, 1 0 ↓X1
[15]   W←(¯1↓Y11),(¯1↓Y21),(¯1↓Y31),(¯1↓YL1),(¯1↓YL2),(2 0 ↓Z
       ), 1 0 ↓ ¯1 0 ↓ 0 ¯1 ↓Z
[16]   Y1H←W+.×(1↓Y11)⌹W
[17]   Y2H←W+.×(1↓Y21)⌹W
[18]   Y3H←W+.×(1↓Y31)⌹W
[19]   W1←(1↓Y2H),(1↓Y3H),(¯1↓Y1H), 3 0 ↓X1
[20]   W2←(1↓Y1H),(¯1↓Y2H), 3 0 ↓X2
[21]   Z1←(2↓Y21),(2↓Y31),(2↓YL1), 3 0 ↓X1
[22]   Z2←(2↓Y11),(2↓YL2), 3 0 ↓X2
[23]   B1←(⌹((⍉W1)+.×Z1))+.×(⍉W1)+.×2↓Y11
[24]   B2←(⌹((⍉W2)+.×Z2))+.×(⍉W2)+.×2↓Y21
[25]   B3←(1↓Y31)⌹(Y2H, 2 0 ↓X3)
[26]   BB← 3 3 ⍴1,(-B1[1]),(-B1[2]),(-B2[1]),1,0,0,(-B3[1]),1
[27]   Q1← 1 7 ⍴B1[4],0,0,B1[5],0,0,B1[6]
[28]   Q2← 1 7 ⍴0,B2[3],0,B2[4],0,B2[5],B2[6]
[29]   Q3← 1 7 ⍴0,B3[2],B3[3],0,B3[4],0,B3[5]
[30]   Q←Q1,[1] Q2,[1] Q3
[31]   P← 3 3 ⍴B1[3],0,0,0,B2[2],0,0,0,0
[32]   R1←(1↓E)⌹¯1↓E←Y11-(Y21,Y31,YL1, 1 0 ↓X1)+.×B1
[33]   R2←(1↓E)⌹¯1↓E←Y21-(Y11,YL2, 1 0 ↓X2)+.×B2
[34]   R3←(1↓E)⌹¯1↓E←Y31-(Y21, 1 0 ↓X3)+.×B3
[35]   RR← 3 3 ⍴R1,0,0,0,R2,0,0,0,R3
[36]   P1←(⌹BB)+.×(P+RR+.×BB)
[37]   P2←(⌹BB)+.×RR+.×P
[38]   P3←(⌹BB)+.×Q
[39]   P4←(⌹BB)+.×RR+.×Q
[40]   YY←⍉(3 31 ⍴Y1,Y2,Y3)
[41]   Y2H←(P1[2;]+.×⍉(1 0 ↓ ¯1 0 ↓YY))-(P2[2;]+.×⍉(¯2 0 ↓YY)
       )+(P3[2;]+.×⍉(2 0 ↓Z))-(P4[2;]+.×⍉(1 0 ↓ ¯1 0 ↓Z))
[42]   Y3H←(P1[3;]+.×⍉(1 0 ↓ ¯1 0 ↓YY))-(P2[3;]+.×⍉(¯2 0 ↓YY)
       )+(P3[3;]+.×⍉(2 0 ↓Z))-(P4[3;]+.×⍉(1 0 ↓ ¯1 0 ↓Z))
[43]   UH1←Y11-(Y21,Y31,XI)+.×B1
[44]   Y2HR←Y2H-R1×¯1↓Y21
[45]   Y3HR←Y3H-R1×¯1↓Y31
[46]   Y1R←(1↓Y11)-R1×¯1↓Y11
[47]   Y2R←(1↓Y21)-R1×¯1↓Y21
[48]   Y3R←(1↓Y31)-R1×¯1↓Y31
[49]   XIR←(1 0 ↓XI)-R1× ¯1 0 ↓XI
[50]   WH←Y2HR,Y3HR,XIR,¯1↓UH1
```

```
[51]  ZR←Y2R,Y3R,XIR,¯1↓UH1
[52]  BR←(⌹((⍉WH)+.×ZR))+.×(⍉WH)+.×Y1R
[53]  AHB←AHB,[1] BR+(0,0,0,0,0,0,R1)
[54]  URH←Y1R-(Y2R,Y3R,XIR)+.×¯1↓BR
[55]  S2←(+/URH*2)÷24
[56]  V←S2×⌹(⍉WH)+.×ZR
[57]  AVHB←AVHB,[1] V[1;1],V[2;2],V[3;3],V[4;4],V[5;5],V[6;6
      ],V[7;7]
[58] ⍝ ****CALCULATION OF HATANAKA C
[59]  URH1←(1↓UH1)-R1×¯1↓UH1
[60]  BC←URH1⌹WH
[61]  AHC←AHC,[1] BC+(B1,R1)
[62]  URH←Y1R-(Y2R,Y3R,XIR)+.×¯1↓BC+(B1,R1)
[63]  S2←(+/URH*2)÷24
[64]  V←S2×⌹(⍉WH)+.×WH
[65]  AVHC←AVHC,[1] V[1;1],V[2;2],V[3;3],V[4;4],V[5;5],V[6;6
      ],V[7;7]
[66]  →L1×⍳(100≥II←II+1)
[67]  '***END'
    ∇
```

APPENDIX III: THE RELATIVE IMPORTANCE OF THE FIRST OBSERVATION IN SIMULTANEOUS EQUATION MODELS

Review of the Findings in the Single Equation Case

Taylor(1981) gave analytical reasons for the differences that exist in the performance of a single equation estimator in the presence of autocorrelation when different specifications for the explanatory variables are used. In general he showed that the specification of the process that generates the explanatory variable affects the performance of different estimators. As was the case in Rao and Griliches(1969), Maeshiro(1976,1979), and Park and Mitchell(1980), he considered the following model

$$y_t = a_1 + b_1 X_t + u_t \qquad t=1,...,T \tag{1}$$

$$u_t = r u_{t-1} + \epsilon_t$$

Following Maeshiro, Rao and Griliches and Park and Mitchell the X variable followed one of the following processes:

1-Stationary autoregressive process of the form

$$X_t = \lambda X_{t-1} + V_t \tag{2}$$

where

$$\mathrm{Var}(X_t) = \mathrm{plim}\ (1/T)(\Sigma X_t^2) = \sigma_x^2 = \sigma^2/(1-\lambda^2)$$

2- Non-stochastic autoregressive process of the form:

$$X_t = \lambda X_{t-1} \qquad t=2,...,T$$

$$X_1 = c_1$$

where $\mathrm{Var}(X_t) = \mathrm{plim}\ (1/T)(\Sigma X_t^2) = \sigma_x^2 = 0$

He showed that the ratio of the variance of the coefficient b when it is estimated by the Cochrane-Orcutt method (b), which omits the first observation, to the GLS (which retains

the first observation) (b_s) is equal to

$$\frac{V(\hat{b}_c)}{V(\hat{b}_s)} = \frac{(1-2\lambda r+r^2)(T-1)\sigma_x^2+(1-r^2)\sigma_x^2+c_1^2(1-r\lambda)^2/(1-\lambda)}{(1-2\lambda r+r^2)(T-1)\sigma_x^2+c_1^2(r-\lambda)^2/(1-\lambda^2)}$$

which goes to one as T approaches infinity if σ_x^2 is nonzero. However, if σ_x^2 approaches zero, we shall have

$$\frac{V(\hat{b}_c)}{V(b_s)} \simeq \frac{(1-r\lambda)^2}{(r-\lambda)^2} = 1 + \frac{(1-r^2)(1-\lambda^2)}{(r-\lambda)^2} \tag{3}$$

Which tends to infinity as r approaches λ.

He also showed that if $\sigma_x^2 = 0$, then

$$V(b_s)=\sigma^2/[(1-r^2)X_1^2+(1-r/\lambda)^2\sum_2^T X_t^2] \tag{4}$$

which suggest that as r tends toward λ, the relative importance of the first observation increases and when $r=\lambda$, the only relevant observation is the first one.

Transforming model (1) so that it is free of autocorrelation and then estimating the resultant model with OLS gives:

$$\sqrt{1-r^2}\, y_1 = b_1\sqrt{1-r^2}\, X_1 + \epsilon_1$$

$$y_t - ry_{t-1} = b_1(X_t - rX_{t-1}) + \epsilon_t \qquad t=2,\ldots T$$

if all observations are utilized;

or

$$\bar{y} = b_1\bar{X} + \epsilon_t$$

This transformation does not necessarily decrease the variability of the X matrix. Since

$$V(\bar{X}_t)/V(X_t) = 1 - 2r\lambda + r^2 \tag{5}$$

the above transformation increases the variance of the exogenous variable whenever ($r>0$, $r>2\lambda$) or ($r<0$, $r<2\lambda$)

and thus increases the precision of the associated estimator of b_1. If X is non-stochastic ($\sigma_x^2=0$) and r becomes equal to λ, we shall have

$$\sqrt{1-r^2}\ y_1 = b\sqrt{1-r^2}\ X_1 + \epsilon_1$$

$$y_t - ry_{t-1} = b_1(X_t - rX_{t-1}) + \epsilon_t$$

or

$$\sqrt{1-r^2}\ y_1 = b_1\sqrt{1-r^2}\ X_1 + \epsilon_1$$

$$y_t - ry_{t-1} = 0 + \epsilon_t$$

which means that as a result of the above transformation, if X is non-stochastic trended ($\sigma_x^2=0$ or $V_t=0$) the only relevant observation is the first one and the Cochrane-Orcutt estimator which omits the first observation, loses all relevant information about the data and this is the reason why its variance tends to infinity (see formula (3)). However, this would not have happened if X process was stochastic, i.e., $\overline{X}_t = V_t$, $t=2,\ldots,T$.

Therefore, Taylor concluded that (1981, p.78):

> Thus Park and Mitchell's (and Maeshiro's) results unambiguously apply to non-stochastic trended variables and to stochastic processes whose realization happen to be trended. On the other hand, Rao and Griliches results pertain to a stochastic exogenous process....

Simultanous Equation Case

The following points may help to explain why the presence of trend in the explanatory variables affects the performance of the simultaneous equation estimators differently from that of the single equation methods. In

other words, why the estimators that employ all T observations do not necessarily perform better than those which utilize T-1 observations when some of the explanatory variables are trended:

1. Unlike the single equation models, simultaneous equation models include some endogenous variables among their explanatory variables. This means that even if all the purely exogenous variables are trended and non-stochastic, the explanatory variables, nevertheless contain some stochastic elements.
2. Many of the limited information methods do not use transformed X's as instruments in their first step. Therefore, since they do not transform the X's, they will not be subject to any possible reduction in the variability of the exogenous variables caused by the autoregressive transformation. Moreover, some of the limited information estimators (like Fair) use the lagged value of the endogenous variables as instruments in the first step. Thus even if the X's are non-stochastic and trended, the Y's are not. Hence, the instruments in the first step estimation will inevitably contain stochastic variables.
3. If some of the X's are non-stochastic and trended and it happens that their trend coefficients are close to the autocorrelation coefficient of the structural equation under consideration, the value of the first observation of the trended X's decreases as the number of the X's

that are not trended increases or the number for which the trend coefficients are not equal to r, increases. This means that even those estimators, i.e., Theil type, that rely on the transformed X's will perform more or less the same whether they use T or T-1 observations if the number of X's are large. This explains the small difference that exist between Theil and M2SLS when the X's are trended. This can also explain the observed improvement in the performance of the Fair estimator which uses non-transformed X's and Y's as instrument relative to the Theil estimator when X_3 and X_6 were trended.

4. It has to be noted that all the experiments conducted in the single equation context have considered only one explanatory variable. It would seem reasonable to expect that as the number of explanatory variable increases (especially if some of them are not trended or their trend coefficients are not close to the autocorrelation coefficient of that equation) the value of the first observation of the trended variable decreases. Hence, the gap between the Cochrane-Orcutt and GLS estimators decreases.

Therefore, the conclusion reached by Taylor that "the case in which the first observation is asymptotically important requires an unusual X process; for typical economic variables, the asymmetry disappears in expectation" seems to be valid for the simultaneous equation system as

well. However, in empirical work, it is advisable to use estimators which employ all T observations as our experiments show.

To test the empirical validity of the proposition outlined above concerning the effect of having more than one explanatory variable in the model on the performance of single equation estimators we conducted the following sampling experiments:

Experiment A: The first experiment was conducted using a model similar to that of Maeshiro and Park and Mitchell, with the following stochastic characteristics:

$$y_t = a_1 + b_1 X_{1t} + u_t$$

$$u_t = r\, u_{t-1} + \epsilon_t$$

$$X_{1t} = \lambda\, X_{1,t-1}$$

$$\epsilon_t \sim IIN(0,\ 0.025)$$

$$r = \lambda = 0.6$$

$$a_1 = b_1 = 1$$

100 samples of 20 observations were created. Then the performance of OLS, CORC, and GLS were examined. Table A3-1 shows the relative RMSE of these estimators. The results clearly support the findings of Maeshiro and Park and Mitchell suggesting the extreme inefficiency of CORC relative to OLS and GLS.

Experiment B: The second experiment was conducted employing the same model as above except adding two regressors to give

TABLE A3-1: Relative Root Mean Squares Errors in the Single Equation Experiment

Coefficient	OLS/CORC	CORC/GLS	OLS/GLS
a_1	0.625	1.677	1.048
b_1	0.008	157.479	1.302

TABLE A3-2: Relative Root Mean Squares Errors in the Single Equation Experiment

Coefficient	OLS/CORC	CORC/GLS	OLS/GLS
a_1	3.33	0.52	1.75
b_1	0.77	1.16	0.90
b_2	3.66	0.38	1.41
b_3	2.07	4.33	8.97

$$y_t = a_1 + b_1X_1 + b_2X_2 + b_3X_3 + u_t$$

where

$$X_{2t} = \lambda_2 X_{2,t-1} + V_{2t}$$

$$X_{3t} = \lambda_3 X_{3,t-1} + V_{3t}$$

$$V_{2t} \sim IIN(0, 0.06)$$

$$V_{3t} \sim IIN(0, 0.09)$$

$$\lambda_2 = 0.2 , \lambda_3 = 0.3$$

The stochastic regressor was added on the basis of the argument that it is unrealistic to assume all the regressors of an equation are non-stochastic.

Table A3-2 shows the relative RMSE of OLS, CORC, and GLS. OLS and GLS which employ all T observations are still performing better on the estimation of the trended variable X_1, but the extreme asymmetry due to the first observation has disappeared. CORC is no longer inferior to OLS. This result shows that the relative asymmetry of the first observation of the non-stochastic trended variable decreases as the number of stochastic regressors increases. Our result supports findings of Rao and Griliches that correcting for autocorrelation results in a gain of efficiency especially when r>0.3. This results also clarifies our findings in the simultaneous equation system that there does not exist a dramatic difference between the limited information estimators which use T and those that employ T-1 observations when the exogenous variables are trended or non-trended.

APPENDIX IV: MONTE CARLO RESULTS CONCERNING THE AUTOREGRESSIVE MODELS WITHOUT LAGGED ENDOGENOUS VARIABLES

This appendix presents the detail statistics on the results of the Monte Carlo experiments using the static model presented in the first part of this thesis. The structural equation under consideration can be written as

$$y_1 = 32 + 0.64\ y_2 + 0.22\ y_3 + 0.65\ X_1 + 0.52\ X_4 + u_1$$

In the following tables we have used normalized (RMSE), normalized trace(MSE), and normalized det(MSE) indices. To define these indices, we first specify the following indices

1- The root mean square error(RMSE) of the jth coefficient is

$$\mathrm{RMSE}(j) = \left[\sum_{i=1}^{N}(\hat{\beta}_{ij} - \beta_j)^2 \ / \ N\right]^{1/2}$$

2- The mean square error matrix is

$$\mathrm{MSE} = \left[\Sigma(\hat{\beta}_i - \beta)(\hat{\beta}_i - \beta)'\right] / \ N$$

3- The trace(MSE) is

$$\mathrm{Trace(MSE)} = \sum_{j=1}^{K} \mathrm{diag(MSE)}$$

Then we can define the normalized indices are as follows

$$\mathrm{Normalized(RMSE_s)} = \mathrm{RMSE_s} / \ \mathrm{RMSE(Theil\ True)}$$

$$\mathrm{Normalized\ tr(MSE_s)} = \mathrm{tr(MSE_s)} \ / \ \mathrm{tr[MSE(Theil\ True)]}$$

$$\mathrm{Normalized\ det(MSE)} = \mathrm{det(MSE_s)} \ / \ \mathrm{det[MSE(Theil\ True)]}$$

TABLE A4-1: Estimate of Bias For the Autoregressive Static Model
(r=0.2, T=30)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	-4.34	-5.20	-26.50	4.07	-1.45	-5.38
B_1	0.14	0.14	0.27	0.17	0.05	0.15
B_2	-0.16	-0.15	-0.06	-0.32	-0.06	-0.16
C_1	-0.12	-0.13	-0.19	-0.17	-0.04	-0.13
C_2	-0.26	-0.28	-0.69	-0.25	-0.11	-0.28
r	-0.08	-0.10	-0.17	-0.12	-0.14	N.A
Aggregate* Bias	0.36	0.38	0.79	0.49	0.20	0.38

*Note that the aggregate bias index does not include the intercept term. In addition the aggregate bias of the 2SLS estimator is on the basis of the four estimated structural coefficients' biases.

TABLE A4-2: Estimate of Bias For the Autoregressive Static Model
(r=0.6, T=30)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	-3.13	-3.32	-16.74	5.74	-4.67	-5.51
B_1	0.11	0.10	0.21	-0.12	0.08	0.16
B_2	-0.13	-0.11	-0.10	0.11	-0.06	-0.18
C_1	-0.09	-0.09	-0.14	0.04	-0.07	-0.13
C_2	-0.21	-0.21	-0.51	0.25	-0.19	-0.31
r	-0.12	-0.15	-0.21	-0.15	-0.27	N.A
Aggregate* Bias	0.31	0.31	0.61	0.34	0.35	0.41

*Note that the aggregate bias index does not include the intercept term. In addition the aggregate bias of the 2SLS estimator is on the basis of the four estimated structural coefficients' biases.

TABLE A4-3: Estimate of Bias For the Autoregressive Static Model ($r=0.9$, $T=30$)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	-2.40	-2.23	-20.18	-9.49	-2.48	-7.06
B_1	0.08	0.06	0.48	0.06	0.04	0.21
B_2	-0.09	-0.06	-0.47	0.04	-0.03	-0.23
C_1	-0.06	-0.05	-0.40	-0.03	0.00	-0.21
C_2	-0.14	-0.12	-1.01	-0.18	-0.10	-0.38
r	-0.15	-0.19	-0.64	-0.31	-0.30	N.A
Aggregate* Bias	0.24	0.24	1.43	0.37	0.32	0.53

*Note that the aggregate bias index does not include the intercept term. In addition the aggregate bias of the 2SLS estimator is on the basis of the four estimated structural coefficients' biases.

TABLE A4-4: Estimate of Bias For the Autoregressive Static Model
(r=0.2, T=60)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	-1.80	-1.83	-11.92	3.71	0.16	-1.99
B_1	0.04	0.04	0.29	-0.09	0.01	0.04
B_2	-0.04	-0.04	-0.30	0.09	-0.01	-0.04
C_1	-0.04	-0.04	-0.27	0.11	-0.02	-0.04
C_2	-0.07	-0.07	-0.51	0.15	-0.02	-0.08
r	-0.03	-0.03	0.03	-0.05	-0.03	N.A
Aggregate* Bias	0.10	0.10	0.71	0.20	0.04	0.08

*Note that the aggregate bias index does not include the intercept term. in addition the aggregate bias of the 2SLS estimator is on the basis of the four estimated structural coefficients' biases.

TABLE A4-5: Estimate of Bias For the Autoregressive Static Model
(r=0.6, T=60)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	-1.46	-1.51	-5.15	1.47	-0.61	-2.80
B_1	0.03	0.04	0.24	-0.02	0.02	0.07
B_2	-0.03	-0.04	-0.31	0.01	-0.02	-0.07
C_1	-0.04	-0.04	-0.23	0.02	-0.04	-0.08
C_2	-0.06	-0.06	-0.42	0.04	-0.04	-0.13
r	-0.02	-0.03	-0.14	-0.05	-0.11	N.A
Aggregate* Bias	0.09	0.10	0.63	0.07	0.13	0.18

*Note that the aggregate bias index does not include the intercept term. In addition the aggregate bias of the 2SLS estimator is on the basis of the four estimated structural coefficients' biases.

TABLE A4-6: Estimate of Bias For the Autoregressive Static Model
($r=0.9$, $T=60$)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	-0.71	0.71	-23.50	0.71	-0.25	-3.37
B_1	0.02	0.02	0.73	0.00	0.00	0.05
B_2	-0.02	-0.02	-0.84	0.00	0.00	-0.04
C_1	-0.02	-0.02	-0.71	0.00	0.00	-0.01
C_2	-0.03	-0.03	-1.27	0.01	0.01	-0.13
r	-0.04	-0.06	-0.67	-0.06	-0.09	N.A
Aggregate* Bias	0.06	0.07	1.95	0.06	0.09	0.15

*Note that the aggregate bias index does not include the intercept term. In addition the aggregate bias of the 2SLS estimator is on the basis of the four estimated structural coefficients' biases.

TABLE A4-7: Normalized Trace(RMSE) Using Autoregressive Static Model
(r=0.2, T=30)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS	True Value
C	1.01	1.10	4.01	5.80	1.58	1.11	7.590
B_1	1.06	1.06	1.61	15.66	1.67	1.09	0.176
B_2	1.14	1.09	1.33	20.34	2.29	1.07	0.207
C_1	1.06	1.06	1.41	13.63	1.82	1.09	0.170
C_2	1.06	1.12	2.18	13.19	1.67	1.09	0.330
Normalized* Det(MSE)	4.04	5.74	161.81	3425781	731.17	1.49	0.377
Normalized* Trace(MSE)	1.14	1.24	3.60	244.99	3.46	1.20	0.212

*Note that this index does not include the intercept term.

TABLE A4-8: Normalized Trace(RMSE) Using Autoregressive Static Model
(r=0.6, T=30)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS	True Value
C	1.04	1.21	3.80	4.38	1.70	1.45	6.706
B_1	1.00	1.02	1.59	3.71	1.76	1.23	0.169
B_2	1.04	1.00	1.09	3.27	1.95	1.14	0.223
C_1	1.04	1.14	1.36	3.64	2.43	1.29	0.136
C_2	1.00	1.15	1.94	4.44	1.79	1.26	0.342
Normalized* Det(MSE)	1.02	4.01	39.07	11669	275.84	2.23	0.094
Normalized* Trace(MSE)	1.02	1.21	3.57	16.10	3.67	1.53	0.210

*Note that this index does not include the intercept term.

TABLE A4-9: Normalized Trace(RMSE) Using Autoregressive Static Model
($r=0.9$, $T=30$)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS	True Value
C	1.04	1.10	6.22	5.98	2.05	2.13	5.690
B_1	1.25	1.08	4.67	3.50	2.33	2.38	0.120
B_2	1.31	1.12	4.56	4.44	2.06	2.20	0.165
C_1	1.30	1.30	6.00	4.10	4.40	2.58	0.104
C_2	1.27	1.14	5.23	3.41	2.50	2.33	0.224
Normalized* Det(MSE)	2.13	2.42	14483.52	55681.32	277.25	154.13	0.091
Normalized* Trace(MSE)	1.64	1.30	25.64	14.35	6.96	5.42	0.103

*Note that this index does not include the intercept term.

TABLE A4-10: Normalized Trace(RMSE) Using Autoregressive Static Model
(r=0.2, T=60)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS	True Value
C	1.02	1.04	3.81	4.45	0.89	1.02	3.871
B_1	1.01	1.03	4.82	5.51	0.92	1.06	0.073
B_2	0.99	1.01	5.59	6.32	1.24	1.07	0.068
C_1	1.00	1.03	4.31	8.62	1.17	1.06	0.077
C_2	1.02	1.05	4.64	4.64	0.96	1.08	0.134
Normalized* Det(MSE)	1.26	1.58	4463.2	10031.6	40.42	1.20	95x10
Normalized* Trace(MSE)	1.02	1.08	22.49	34.55	1.08	1.15	0.034

*Note that this index does not include the intercept term.

TABLE A4-11: Normalized Trace(RMSE) Using Autoregressive Static Model
(r=0.6, T=60)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS	True Value
C	1.03	1.03	4.68	1.62	1.26	1.80	2.719
B_1	1.02	1.05	6.09	1.35	1.33	1.96	0.056
B_2	1.02	1.05	7.86	1.57	1.47	2.07	0.058
C_1	1.03	1.06	5.25	1.20	1.73	2.44	0.064
C_2	1.02	1.05	5.59	1.32	1.35	1.94	0.104
Normalized* Det(MSE)	1.39	1.58	33804.35	11.48	79.35	136.96	46*10
Normalized* Trace(MSE)	1.01	1.14	36.90	1.85	2.13	4.36	0.021

*Note that this index does not include the intercept term.

TABLE A4-12: Normalized Trace(RMSE) Using Autoregressive Static Model
(r=0.9, T=60)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS	True Value
C	1.00	1.00	11.99	1.88	1.31	3.47	2.091
B_1	1.00	1.02	19.53	1.68	1.29	4.29	0.038
B_2	0.99	1.01	19.66	1.76	1.21	4.21	0.044
C_1	1.00	1.02	15.63	1.39	1.59	5.07	0.046
C_2	1.00	1.03	18.40	1.50	1.13	3.97	-0.070
Normalized* Det(MSE)	0.97	1.03	159190.7	17.05	12.77	46184.97	17×10
Normalized* Trace(MSE)	1.00	1.12	347.49	2.50	1.63	19.22	0.010

*Note that this index does not include the intercept term.

APPENDIX V: MONTE CARLO RESULTS OF THE DYNAMIC AUTOREGRESSIVE MODELS

This appendix presents detail statistics of the result of the Monte Carlo experiments using the dynamic autoregressive model outlined in chapter six.

The structural equation under consideration can be presented as

$y_1 = 32 + 0.64\ y_2 + 0.22\ y_3 + 0.25\ y_{1,-1} + 0.65\ X_1 + 0.52\ X_4 + u_1$

TABLE A5-1: Estimated Bias on the Basis of the Mean Using Dynamic Autoregressive Model
(r=0.2, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	-20.71	-18.13	-15.16	69.44	4.85	-17.61	0.65
B	0.27	0.32	0.26	0.00	0.27	0.26	0.12
B	-0.29	-0.45	-0.28	-0.05	-0.24	-0.28	-0.14
B	-0.20	-0.22	-0.20	-0.04	-0.23	-0.20	-0.09
C	-0.07	-0.04	-0.05	0.17	0.04	-0.06	-0.02
C	-0.51	-0.54	-0.46	0.17	-0.34	-0.49	-0.19
r	0.09	0.01	0.00	-0.08	0.08	0.06	N.A
Aggregate* Bias	0.69	0.80	0.63	0.26	0.55	0.66	0.28

*Note that this index does not include the intercept term.

TABLE A5-2: Estimated Bias on the Basis of the Mean Using Dynamic Autoregressive Model
(r=0.6, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	-8.20	4.36	-5.82	23.34	3.92	-7.89	-4.37
B	0.21	0.40	0.20	0.67	0.26	0.21	0.11
B	-0.22	-0.64	-0.21	-1.18	-0.27	-0.21	-0.13
B	-0.16	-0.28	-0.16	-0.45	-0.21	-0.16	-0.08
C	-0.05	-0.06	-0.04	0.00	0.04	-0.05	-0.01
C	-0.41	-0.60	-0.38	-1.58	-0.36	-0.40	-0.15
r	-0.08	-0.29	-0.18	-0.12	-0.18	-0.11	N.A
Aggregate* Bias	0.54	1.05	0.54	2.10	0.59	0.54	0.24

*Note that this index does not include the intercept term.

TABLE A5-3: Estimated Bias on the Basis of the Mean Using Dynamic Autoregressive Model
(r=0.9, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	-33.71	-30.47	-43.47	-51.81	-44.29	-38.60	-45.34
B	0.19	0.56	0.26	-0.01	0.46	0.20	0.31
B	-0.21	-0.83	-0.27	0.08	-0.49	-0.21	-0.36
B	-0.13	-0.38	-0.17	0.04	-0.34	-0.13	-0.21
C	-0.04	-0.05	-0.04	0.04	0.00	-0.04	-0.02
C	-0.39	-0.92	-0.49	0.04	-0.68	-0.40	-0.52
r	-0.20	-0.56	-0.34	-0.16	-0.40	-0.22	N.A
Aggregate* Bias	0.54	1.52	0.73	0.19	1.09	0.56	0.74

*Note that this index does not include the intercept term.

TABLE A5-4: Estimated Bias on the Basis of the Mean Using Dynamic Autoregressive Model
(r=0.2, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	6.68	-1.62	6.13	-37.45	15.39	6.44	1.33
B	0.17	0.17	0.17	1.18	0.27	0.17	0.05
B	-0.17	-0.07	-0.17	-1.24	-0.26	-0.17	-0.06
B	-0.14	-0.15	-0.14	-0.90	-0.22	-0.14	-0.04
C	-0.13	-0.17	-0.12	-0.17	-0.03	-0.13	-0.03
C	-0.30	-0.32	-0.29	-1.98	-0.30	-0.30	-0.09
r	0.07	0.10	0.03	0.60	0.13	0.06	N.A
Aggregate* Bias	0.43	0.44	0.42	3.06	0.44	0.43	0.13

*Note that this index does not include the intercept term.

TABLE A5-5: Estimated Bias on the Basis of the Mean Using Dynamic Autoregressive Model
(r=0.6, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	7.27	4.74	2.55	2.02	13.83	7.33	-9.99
B	0.12	0.25	0.12	0.11	0.26	0.12	-0.02
B	-0.11	-0.40	-0.11	-0.11	-0.25	-0.10	-0.01
B	-0.11	-0.17	-0.10	-0.09	-0.22	-0.10	0.03
C	-0.12	-0.17	-0.11	-0.07	-0.06	-0.12	0.02
C	0.23	-0.32	-0.22	-0.22	-0.34	-0.22	0.07
r	0.00	-0.18	-0.05	-0.11	0.00	0.00	N.A
Aggregate* Bias	0.33	0.64	0.32	0.31	0.55	0.31	0.08

*Note that this index does not include the intercept term.

TABLE A5-6: Estimated Bias on the Basis of the Mean Using Dynamic Autoregressive Model
(r=0.9, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	-8.41	-31.93	-36.45	-23.84	-30.02	-10.34	-34.61
B	0.11	0.60	0.18	0.15	0.53	0.10	0.43
B	-0.10	-0.74	-0.17	-0.14	-0.55	-0.09	-0.55
B	-0.08	-0.44	-0.13	-0.11	-0.40	-0.08	-0.30
C	-0.11	-0.50	-0.19	-0.09	-0.28	-0.10	-0.26
C	-0.19	-0.88	-0.29	-0.32	-0.82	-0.18	-0.63
r	-0.07	-0.47	-0.21	-0.23	-0.30	-0.07	N.A
Aggregate* Bias	0.29	1.53	0.49	0.47	1.26	0.27	1.02

*Note that this index does not include the intercept term.

TABLE A5-7: Estimated Bias on the Basis of the Median Using Dynamic Autoregressive Model
(r=0.2, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	-19.51	-8.32	-16.69	15.66	2.67	-16.52	0.86
B	0.27	0.32	0.26	-0.05	0.26	0.26	0.12
B	-0.28	-0.43	-0.27	0.03	-0.26	-0.28	-0.14
B	-0.21	-0.34	-0.20	0.03	-0.24	-0.20	-0.09
C	-0.07	-0.04	-0.05	0.03	0.03	-0.06	-0.02
C	-0.53	-0.53	-0.47	0.10	-0.36	-0.49	-0.18
r	0.10	0.00	-0.02	-0.02	-0.18	0.08	N.A
Aggregate* Bias	0.70	0.83	0.64	0.22	0.57	0.66	0.27

*Note that this index does not include the intercept term.

TABLE A5-8: Estimated Bias on the Basis of the Median Using Dynamic Autoregressive Model
(r=0.6, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	-6.12	6.59	-3.78	8.09	6.47	-4.55	-2.76
B	0.22	0.38	0.20	0.10	0.28	0.21	0.10
B	-0.22	-0.54	-0.21	-0.14	-0.28	-0.20	-0.12
B	-0.16	-0.26	-0.16	-0.10	-0.22	-0.16	-0.07
C	-0.05	-0.02	-0.04	0.02	0.04	-0.06	-0.01
C	-0.41	-0.57	-0.38	-0.20	-0.39	-0.41	-0.16
r	-0.07	-0.31	-0.17	-0.27	-0.14	-0.09	N.A
Aggregate* Bias	0.55	0.96	0.53	0.39	0.61	0.54	0.23

*Note that this index does not include the intercept term.

TABLE A5-9: Estimated Bias on the Basis of the Median Using Dynamic Autoregressive Model
(r=0.9, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	-39.37	-30.79	-45.33	-46.50	-47.11	-39.75	-46.28
B	0.19	0.55	0.26	-0.01	0.46	0.20	0.29
B	-0.21	-0.80	-0.27	0.04	-0.46	-0.20	-0.37
B	-0.13	-0.38	-0.17	0.03	-0.33	-0.13	-0.19
C	-0.04	-0.04	-0.05	0.00	-0.01	-0.04	-0.02
C	-0.39	-0.90	-0.48	0.01	-0.67	-0.40	-0.50
r	-0.17	-0.53	-0.34	-0.25	-0.38	-0.19	N.A
Aggregate* Bias	0.53	1.48	0.72	0.25	1.06	0.54	0.71

*Note that this index does not include the intercept term.

TABLE A5-10: Estimated Bias on the Basis of the Median Using Dynamic Autoregressive Model
(r=0.2, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	5.51	-0.80	5.66	5.79	14.53	5.54	0.41
B	0.17	0.17	0.17	-0.02	0.28	0.17	0.04
B	-0.18	-0.09	-0.16	0.01	-0.16	-0.18	-0.05
B	-0.14	-0.16	-0.14	0.00	-0.24	-0.15	-0.04
C	-0.13	-0.18	-0.12	0.02	-0.03	-0.13	-0.03
C	-0.29	-0.37	-0.30	0.00	-0.29	-0.30	-0.08
r	0.08	0.12	0.02	-0.05	0.13	0.07	N.A
Aggregate* Bias	0.43	0.50	0.42	0.06	0.51	0.44	0.11

*Note that this index does not include the intercept term.

TABLE A5-11: Estimated Bias on the Basis of the Median Using Dynamic Autoregressive Model
(r=0.6, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	5.75	6.50	2.23	9.87	10.64	5.42	-9.95
B	0.12	0.26	0.12	0.17	0.28	0.12	-0.04
B	-0.11	-0.27	-0.11	-0.19	-0.26	-0.10	0.00
B	-0.11	-0.18	-0.10	-0.13	-0.22	-0.10	-0.03
C	-0.12	-0.18	-0.12	-0.09	-0.06	-0.12	0.01
C	-0.23	-0.35	-0.23	-0.26	-0.34	-0.22	0.06
r	0.02	-0.19	-0.03	-0.08	0.03	0.03	N.A
Aggregate* Bias	0.33	0.57	0.32	0.40	0.45	0.31	0.08

*Note that this index does not include the intercept term.

TABLE A5-12: Estimated Bias on the Basis of the Median Using Dynamic Autoregressive Model
(r=0.9, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS
C	-6.30	-30.49	-36.66	-29.28	-28.13	-9.28	-33.66
B	0.11	0.59	0.17	0.19	0.56	0.10	0.48
B	-0.10	-0.79	-0.16	-0.29	-0.55	-0.09	-0.63
B	-0.09	-0.44	-0.12	-0.11	-0.41	-0.08	-0.34
C	-0.11	-0.54	-0.18	-0.17	-0.29	-0.10	-0.30
C	-0.19	-0.84	-0.28	-0.30	-0.82	-0.17	-0.69
r	-0.05	-0.48	-0.19	-0.24	-0.29	-0.05	N.A
Aggregate* Bias	0.28	1.55	0.46	0.56	1.27	0.26	1.14

*Note that this index does not include the intercept term.

TABLE A5-13: Normalized Root Mean Squares Error Using Dynamic Autoregressive Model
(r=0.2, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	1.07	1.64	0.85	9.90	0.90	0.98	0.71	31.58
B	1.04	1.44	0.96	8.56	1.07	1.00	0.56	0.27
B	1.02	2.00	1.00	12.07	0.93	1.00	0.65	0.29
B	1.05	1.30	1.00	7.40	1.20	1.00	0.60	0.20
C	1.04	1.56	0.90	13.25	1.04	0.90	1.00	0.08
C	1.06	1.34	0.96	7.08	0.76	1.02	0.50	0.50
Normalized* Det(MSE)	2.15	5707.3	1.46	5.4x10	236.10	2.63	4.88	2x10
Normalized* Trace(MSE)	1.10	2.25	0.94	73.82	0.81	1.02	0.31	0.453

*Note that this index does not include the intercept term.

TABLE A5-14: Normalized Root Mean Squares Error Using Dynamic Autoregressive Model
(r=0.6, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	1.10	2.57	0.98	4.46	1.48	0.98	0.95	29.65
B	1.01	3.50	1.00	23.88	1.31	1.01	0.76	0.22
B	1.04	5.71	1.03	36.74	1.38	1.03	0.98	0.22
B	1.01	3.17	0.99	20.71	1.38	0.99	0.71	0.17
C	1.05	6.59	1.00	40.45	3.56	1.03	1.39	0.07
C	0.98	2.19	0.92	33.08	0.98	0.97	0.63	0.44
Normalized* Det(MSE)	3.74	970833	3.08	43x10	240.42	2.49	63.33	24x10
Normalized* Trace(MSE)	0.99	11.11	0.91	1003.49	1.48	0.97	0.54	0.324

*Note that this index does not include the intercept term.

TABLE A5-15: Normalized Root Mean Squares Error Using Dynamic Autoregressive Model
(r=0.9, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	1.14	1.22	1.19	2.43	1.23	1.09	1.23	47.72
B	1.10	3.54	1.43	7.76	2.45	1.13	1.99	0.20
B	1.08	4.98	1.39	10.95	2.38	1.10	2.18	0.22
B	1.07	3.45	1.41	7.66	2.62	1.09	1.93	0.14
C	1.20	4.37	1.61	17.67	1.92	1.33	2.60	0.05
C	1.07	2.77	1.34	5.90	1.90	1.10	1.60	0.40
Normalized* Det(MSE)	1.91	250411	16.14	17x10	3336.08	1.57	1632.62	6x10
Normalized* Trace(MSE)	1.16	11.84	1.88	57.98	4.55	1.22	3.28	0.271

*Note that this index does not include the intercept term.

TABLE A5-16: Normalized Root Mean Squares Error Using Dynamic Autoregressive Model
(r=0.2, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	1.28	1.90	1.23	54.85	2.38	1.25	1.06	8.32
B	1.02	1.18	1.00	33.02	1.63	1.02	0.84	0.18
B	1.02	1.18	1.01	32.85	1.56	1.01	0.65	0.19
B	1.03	1.27	1.01	31.10	1.68	1.03	0.68	0.15
C	1.10	1.62	1.02	47.31	0.46	1.11	0.77	0.13
C	1.03	1.34	1.01	31.73	1.11	1.03	0.52	0.31
Normalized* Det(MSE)	4.25	82x10	5.53	1.33x10	33333.33	4.36	10.77	1.6x10
Normalized* Trace(MSE)	1.09	1.74	1.04	1153.12	1.80	1.09	0.42	0.20

*Note that this index does not include the intercept term.

TABLE A5-17: Normalized Root Mean Squares Error Using Dynamic Autoregressive Model
(r=0.6, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	1.20	2.00	0.95	7.35	1.53	1.22	1.22	15.21
B	0.99	2.39	1.00	11.57	2.09	0.99	1.07	0.13
B	0.99	4.85	0.99	21.62	2.00	0.97	1.32	0.13
B	1.00	1.94	0.99	9.30	2.00	1.00	1.03	0.11
C	1.00	1.94	0.96	9.47	1.52	0.99	0.93	0.13
C	0.99	1.81	0.98	10.69	1.49	0.98	1.05	0.25
Normalized* Det(MSE)	2.87	19x10	3.58	1.1x10	33720.9	2.70	12790.7	1.72x10
Normalized* Trace(MSE)	0.99	6.48	0.98	158.93	4.21	0.98	1.16	0.125

*Note that this index does not include the intercept term.

TABLE A5-18: Normalized Root Mean Squares Error Using Dynamic Autoregressive Model
(r=0.9, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	1.04	2.08	1.77	7.60	1.44	0.97	1.62	24.50
B	1.00	5.74	1.80	38.52	4.51	0.95	4.18	0.12
B	0.98	7.69	1.69	56.31	4.31	0.92	4.85	0.13
B	0.91	5.15	1.62	34.44	4.14	0.90	3.74	0.10
C	0.98	4.72	1.79	21.63	2.44	0.93	2.52	0.12
C	0.97	4.89	1.63	35.02	3.89	0.91	3.48	0.22
Normalized* Det(MSE)	1.25	36x10	653.50	1.09x10	2x10	0.83	25x10	6.58x10
Normalized* Trace(MSE)	0.94	30.14	2.84	1470.29	15.35	0.84	14.11	0.104

*Note that this index does not include the intercept term.

TABLE A5-19: Normalized Interdecile Range Using Dynamic Autoregressive Model
(r=0.2, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	1.15	2.27	0.94	5.11	1.18	1.01	0.91	58.79
B	1.14	3.43	1.15	14.96	1.68	1.17	1.75	0.14
B	1.03	3.85	1.06	13.63	1.62	1.06	1.62	0.18
B	1.21	1.24	1.32	14.62	2.20	1.30	1.78	0.09
C	1.07	2.55	1.12	8.57	1.63	1.10	1.65	0.10
C	1.27	3.48	1.17	14.70	1.52	1.25	1.52	0.26
Normalized* Trace	1.16	3.18	1.15	13.71	1.66	1.18	1.63	0.77

*Note that this index does not include the intercept term.

TABLE A5-20: Normalized Interdecile Range Using Dynamic Autoregressive Model
($r=0.6$, $T=30$)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	1.08	1.58	0.88	2.16	1.07	0.98	0.92	76.61
B	1.13	3.53	1.20	13.67	1.80	1.13	2.07	0.15
B	1.01	4.45	0.96	14.06	1.21	1.00	1.70	0.22
B	1.06	3.51	1.18	11.32	1.84	1.05	2.18	0.11
C	1.30	2.50	1.30	15.70	2.20	1.10	2.00	0.10
C	1.10	2.97	1.13	11.84	1.35	1.13	1.74	0.31
Normalized* Trace	1.11	3.48	1.13	13.18	1.56	1.09	1.87	0.89

*Note that this index does not include the intercept term.

TABLE A5-21: Normalized Interdecile Range Using Dynamic Autoregressive Model
(r=0.9, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	0.96	1.11	0.82	1.67	0.82	0.80	0.88	108.29
B	1.11	5.16	1.58	9.84	2.26	1.16	3.21	0.19
B	1.00	6.23	1.46	10.61	1.73	1.04	2.96	0.26
B	1.14	4.79	1.50	9.43	2.36	1.14	3.29	0.14
C	1.00	4.33	1.56	8.44	2.44	1.11	3.56	0.09
C	1.16	3.71	1.44	8.34	2.05	1.18	2.55	0.38
Normalized* Trace	1.09	4.78	1.49	9.32	2.08	1.13	2.95	1.06

*Note that this index does not include the intercept term.

TABLE A5-22: Normalized Interdecile Range Using Dynamic Autoregressive Model
(r=0.2, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	1.24	1.74	1.26	3.69	1.64	1.27	1.24	18.17
B	1.06	2.22	0.92	10.00	1.94	1.05	1.17	0.14
B	1.02	3.25	1.02	10.86	1.96	1.04	1.20	0.16
B	1.15	2.29	1.08	10.00	2.11	1.10	1.29	0.11
C	1.34	2.73	1.21	9.82	1.09	1.32	1.16	0.11
C	1.10	2.63	1.03	10.86	1.42	1.16	1.13	0.23
Normalized* Trace	1.12	2.65	1.04	10.42	1.69	1.13	1.18	0.76

*Note that this index does not include the intercept term.

TABLE A5-23: Normalized Interdecile Range Using Dynamic Autoregressive Model
(r=0.6, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	1.16	2.38	0.99	3.36	1.37	1.14	1.13	34.20
B	0.93	3.40	1.00	14.80	1.47	1.00	1.87	0.15
B	0.94	6.11	0.94	21.22	1.00	0.96	1.78	0.18
B	1.00	3.00	1.18	14.27	1.91	1.18	2.09	0.11
C	1.11	2.88	1.21	11.21	1.44	1.16	1.67	0.13
C	1.00	2.89	1.08	12.93	1.33	1.00	1.89	0.25
Normalized* Trace	1.00	3.70	1.07	14.99	1.38	1.04	1.85	0.82

*Note that this index does not include the intercept term.

TABLE A5-24: Normalized Interdecile Range Using Dynamic Autoregressive Model
(r=0.9, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(B)	Hatanaka(C)	Dhrymes	2SLS	True Value
C	1.01	1.55	1.03	1.85	0.79	0.85	0.73	59.79
B	1.01	6.12	2.09	16.19	1.94	1.05	5.11	0.14
B	1.03	7.62	1.66	18.60	1.55	1.09	4.40	0.19
B	1.00	5.45	2.09	13.82	2.18	1.00	5.36	0.11
C	0.96	4.44	1.76	9.28	1.90	0.92	2.88	0.15
C	0.99	4.98	1.94	14.94	2.47	0.99	4.71	0.26
Normalized* Trace	1.00	5.72	1.89	14.81	1.04	1.01	4.46	0.86

*Note that this index does not include the intercept term.

APPENDIX VI: MONTE CARLO RESULTS OF TESTS OF HYPOTHESES FOR THE STATIC AUTOREGRESSIVE MODEL

This appendix presents the detail statistics concerning the Type I error and the power of the test of significance for the autoregressive model without lagged endogenous variables.

TABLE A6-1: Percentage of Type I Error at 5% Significant Level Using Static Model (r=0.2, T=30)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	13	18	50	24	15	16
B	32	30	50	46	25	33
B	24	24	6	39	21	22
C	21	21	18	18	11	22
C	31	31	66	28	20	33
Average*	27	26	35	33	19	27

*Note that this index does not include the intercept term.

TABLE A6-2: Percentage of Type I Error at 5% Significant Level Using Static Model
(r=0.6, T=30)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	10	11	41	25	14	7
B	21	22	53	40	26	36
B	18	14	23	22	16	20
C	12	11	10	12	12	15
C	22	22	54	29	26	30
Average*	18	17	35	26	20	25

*Note that this index does not include the intercept term.

TABLE A6-3: Percentage of Type I Error at 5% Significant Level Using Static Model (r=0.9, T=30)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	10	8	21	32	10	9
B	11	13	59	37	18	36
B	7	9	21	28	11	23
C	6	7	19	14	8	17
C	13	14	59	29	19	28
Average*	9	11	40	27	14	26

*Note that this index does not include the intercept term.

TABLE A6-4: Percentage of Type I Error at 5% Significant Level Using Static Model (r=0.2, T=60)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	9	10	30	6	8	9
B	16	17	54	16	13	15
B	12	12	35	10	12	11
C	11	12	48	5	9	10
C	9	11	51	7	9	11
Average*	12	13	47	10	11	12

*Note that this index does not include the intercept term.

TABLE A6-5: Percentage of Type I Error at 5% Significant Level Using Static Model
(r=0.6, T=60)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	6	7	21	6	6	8
B	12	11	49	10	15	23
B	9	9	40	6	16	21
C	4	3	34	3	11	14
C	6	8	42	6	13	14
Average*	8	8	41	6	14	18

*Note that this index does not include the intercept term.

TABLE A6-6: Percentage of Type I Error at 5% Significant Level Using Static Model (r=0.9, T=60)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	1	1	53	3	4	9
B	6	7	99	9	8	34
B	8	8	71	7	10	25
C	4	4	91	4	7	26
C	4	4	99	7	5	21
Average*	6	6	90	7	8	27

*Note that this index does not include the intercept term.

TABLE A6-7: Percentage Power of the Test of Significance at 5% level Using Static Model
($r=0.2$, $T=30$)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	74	72	6	82	75	72
B	90	92	95	84	83	91
B	3	4	5	41	7	3
C	65	70	70	62	65	67
C	1	1	4	29	6	1
Average*	40	42	44	54	40	41

*Note that this index does not include the intercept term.

TABLE A6-8: Percentage Power of the Test of Significance at 5% level Using Static Model (r=0.6, T=30)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	79	78	28	87	71	60
B	85	87	92	85	86	91
B	4	4	8	27	6	1
C	65	66	70	71	61	55
C	2	3	6	35	9	1
Average*	39	40	44	55	41	37

*Note that this index does not include the intercept term.

TABLE A6-9: Percentage Power of the Test of Significance at 5% level Using Static Model
(r=0.9, T=30)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	72	71	17	74	67	40
B	91	91	93	85	85	89
B	4	4	12	33	6	7
C	64	66	35	71	60	32
C	9	9	33	41	18	3
Average*	42	43	43	58	42	33

*Note that this index does not include the intercept term.

TABLE A6-10: Percentage Power of the Test of Significance at 5% level Using Static Model (r=0.2, T=60)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	100	100	58	100	100	98
B	96	98	89	92	93	96
B	11	11	3	35	15	11
C	97	98	56	98	99	97
C	44	44	7	88	64	41
Average*	62	63	39	78	68	61

*Note that this index does not include the intercept term.

TABLE A6-11: Percentage Power of the Test of Significance at 5% level Using Static Model (r=0.6, T=60)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	99	99	65	100	99	95
B	98	99	92	99	96	94
B	14	10	13	52	16	7
C	96	96	53	98	90	73
C	55	54	8	91	58	25
Average*	66	65	42	85	65	50

*Note that this index does not include the intercept term.

TABLE A6-12: Percentage Power of the Test of Significance at 5% level Using Static Model
(r=0.9, T=60)

Coefficient	Theil	MG2SLS	Fair	ALIML	GLIML	2SLS
C	98	98	15	98	98	71
B	99	99	100	99	100	85
B	23	23	55	51	25	10
C	98	98	6	98	96	41
C	69	70	81	90	69	3
Average*	72	73	61	85	73	35

*Note that this index does not include the intercept term.

APPENDIX VII: MONTE CARLO RESULTS OF TESTS OF HYPOTHESES FOR THE DYNAMIC AUTOREGRESSIVE MODEL

This appendix presents the detail statistics concerning the Type I error and the power of the test of significance for the dynamic autoregressive model studied in the second part of this thesis.

TABLE A7-1: Percentage Type I Error at 5% Level of Significance Using Dynamic Model (r=0.2, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	12	11	18	2	18	9
B	92	47	88	58	92	32
B	72	30	63	16	72	26
B	96	51	92	58	95	34
C	6	3	4	1	5	4
C	81	38	68	15	76	18
Average*	69	34	63	30	68	23

*Note that this index does not include the intercept term.

TABLE A7-2: Percentage Type I Error at 5% Level of Significance Using Dynamic Model
($r=0.6$, $T=30$)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	14	14	20	4	15	23
B	79	47	75	76	80	29
B	54	35	41	41	49	25
B	80	51	76	69	80	29
C	10	7	7	2	10	8
C	56	37	45	18	55	17
Average*	56	35	49	41	55	22

*Note that this index does not include the intercept term.

TABLE A7-3: Percentage Type I Error at 5% Level of Significance Using Dynamic Model
($r=0.9$, $T=30$)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	27	20	44	15	29	56
B	53	40	63	80	60	45
B	31	27	37	54	32	37
B	49	40	53	71	49	46
C	8	5	5	1	9	9
C	40	36	47	40	45	38
Average*	36	30	41	49	39	35

*Note that this index does not include the intercept term.

TABLE A7-4: Percentage Type I Error at 5% Level of Significance Using Dynamic Model
(r=0.2, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	10	2	11	2	10	10
B	74	38	72	68	76	22
B	48	1	44	30	49	13
B	73	48	74	64	75	25
C	38	27	36	4	39	13
C	57	41	52	25	58	15
Average*	58	31	56	38	59	18

*Note that this index does not include the intercept term.

TABLE A7-5: Percentage Type I Error at 5% Level of Significance Using Dynamic Model (r=0.6, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	9	9	13	2	10	28
B	57	40	58	89	58	12
B	28	28	22	68	28	10
B	60	41	58	83	63	12
C	29	22	24	4	27	13
C	35	31	34	36	33	13
Average*	42	32	39	56	42	12

*Note that this index does not include the intercept term.

TABLE A7-6: Percentage Type I Error at 5% Level of Significance Using Dynamic Model
(r=0.9, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	23	28	64	28	20	65
B	34	54	48	99	37	67
B	15	29	29	87	16	63
B	33	50	40	99	34	62
C	22	28	34	23	22	40
C	20	40	35	77	20	58
Average*	25	40	37	77	26	58

*Note that this index does not include the intercept term.

TABLE A7-7: Percentage Power of Test of Significance at 5% Level Using Dynamic Model
(r=0.2, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	11	11	19	4	14	22
B	100	89	100	94	100	95
B	11	5	9	4	10	4
B	18	12	20	8	19	39
C	96	76	97	32	96	84
C	6	7	3	4	6	11
Average*	46	38	46	28	46	47

*Note that this index does not include the intercept term.

TABLE A7-8: Percentage Power of Test of Significance at 5% Level Using Dynamic Model (r=0.6, T=30)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	21	31	34	9	30	37
B	100	92	100	99	100	91
B	8	16	8	11	10	9
B	39	9	32	20	37	42
C	89	68	87	33	88	71
C	5	1	5	5	5	16
Average*	48	37	46	34	48	46

*Note that this index does not include the intercept term.

TABLE A7-9: Percentage Power of Test of Significance at 5% Level Using Dynamic Model
($r=0.9$, $T=30$)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	14	10	23	5	16	29
B	100	85	100	98	100	94
B	3	18	10	26	7	24
B	44	23	24	37	40	26
C	87	50	75	24	86	61
C	7	10	9	11	9	16
Average*	48	37	44	39	48	44

*Note that this index does not include the intercept term.

TABLE A7-10: Percentage Power of Test of Significance at 5% Level Using Dynamic Model
(r=0.2, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	78	46	79	32	80	68
B	100	92	100	100	100	98
B	6	2	5	16	7	7
B	62	6	59	22	63	75
C	100	79	100	50	100	100
C	37	5	36	17	36	51
Average*	61	37	60	41	61	66

*Note that this index does not include the intercept term.

TABLE A7-11: Percentage Power of Test of Significance at 5% Level Using Dynamic Model
(r=0.6, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	54	40	52	27	56	39
B	100	90	100	100	100	90
B	20	10	17	20	19	15
B	84	18	89	22	90	78
C	98	57	98	48	98	93
C	50	8	47	18	51	53
Average*	70	37	70	42	72	66

*Note that this index does not include the intercept term.

TABLE A7-12: Percentage Power of Test of Significance at 5% Level Using Dynamic Model
(r=0.9, T=60)

Coefficient	Theil	Fair	Hatanaka(A)	Hatanaka(C)	Dhrymes	2SLS
C	8	10	27	4	8	28
B	100	93	100	100	100	97
B	14	25	14	59	22	51
B	81	27	50	77	85	50
C	96	21	76	26	95	55
C	44	13	32	26	52	38
Average*	67	36	54	58	71	58

*Note that this index does not include the intercept term.